

**ATTORNEY'S DOCKET NUMBER: 2007651-0001**  
**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

<b>Applicants:</b>	Jack <i>et al.</i>	<b>Examiner:</b>	Hutson, Richard G.
<b>Serial No.:</b>	10/089,027	<b>Art Unit:</b>	1652
<b>Filing Date:</b>	March 26, 2002	<b>Confirmation No.:</b>	9409
<b>Title:</b>	INCORPORATION OF MODIFIED NUCLEOTIDES BY ARCHAEON ANA POLYMERASES AND RELATED METHODS		

Mail Stop Appeal Brief - Patents  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

**ELECTRONICALLY FILED VIA EFS WEB**

Sir:

**AMENDED APPEAL BRIEF UNDER 37 C.F.R. § 41.37**

Appellants hereby appeal to the Board of Patent Appeals and Interferences (the “Board”) from the Examiner’s final rejection of pending claims 32-42 of the above-referenced application.

A final Office Action was mailed on April 21, 2009. A Notice of Appeal was filed on July 21, 2009. An original Appeal Brief was filed on December 23, 2009 with a Petition under 37 C.F.R. § 1.136 for four month extension of time. An electronic payment of the \$270.00 fee for filing an appeal brief under 37 C.F.R. 41.20(b)(2) and \$865 fee for an extension of time was filed concurrently with the original Appeal Brief.

A Notice of Non-Compliant Appeal Brief was mailed on January 25, 2010, setting a deadline of February 25, 2010 for filing an Amended Appeal Brief that complies with the requirements of 37 C.F.R. § 41.37. Thus, the present Amended Appeal Brief is timely filed on February 2, 2010.

Applicant believes that no further petitions and fees are required for this Appeal Brief to be entered. Please consider this a conditional petition for any additional extensions, if needed, and please charge any additional fees or credit any overpayments that may be required to our Deposit Account No. 03-1721 referencing attorney docket number 2007651-0001.

**REAL PARTY IN INTEREST**

As a result of an assignment by the inventors in the present application, the real party in interest in this application is New England Biolabs, Inc. The assignment to New England Biolabs, Inc. was recorded in the Patent and Trademark Office at Reel 012921, Frame 0103.

### **RELATED APPEALS AND INTERFERENCES**

No other appeals or interferences are known to Appellants, Appellants' legal representative, or Appellants' assignee that will directly affect or be directly affected by the Board's decision in this appeal. Similarly, no such appeals or interferences are known that may have a bearing on the Board's decision in this appeal.

### **STATUS OF CLAIMS**

The application was filed with 31 claims. Various claims were amended and/or cancelled in Amendments filed on June 7, 2002, April 18, 2005 (not entered), August 15, 2005, May 4, 2006 (not entered), May 30, 2006 (not entered), July 3, 2006, and February 24, 2007. Pending claims 2-4, 13-22, and 27-31 were canceled in the amendment filed October 31, 2007, and new claims 32-43 were presented. Claim 32 was amended in an Amendment filed on August 29, 2008 (not entered) and an Amendment filed on January 9, 2009. Claims 32-42 were finally rejected in an Office Action mailed April 21, 2009. Claim 43 is objected to for depending on rejected claims 32 and 33.

Thus, claims 1-31 are canceled, claims 32-42 are rejected, and claim 43 is objected to. The rejection of claims 32-42 is hereby appealed. A listing of the claims is provided in the attached **Claims Appendix**.

### **STATUS OF AMENDMENTS**

There are no outstanding amendments to the claims.

## **SUMMARY OF CLAIMED SUBJECT MATTER**

DNA polymerases are enzymes that catalyze polymerization of nucleotides into a DNA strand. The present invention encompasses the finding that a certain class of DNA polymerases has the ability to incorporate a particular modified type of nucleotide, acyclonucleotides, into DNA strands. The present claims therefore recite use of DNA polymerases from that class (defined by the present of a particular amino acid motif whose presence is shown to correlate with the activity) to incorporate acyclonucleotides into a polynucleotide chain.

Independent claim 32 and dependent claims 33-43 specifically recite methods comprising steps of providing a DNA polymerase of the relevant class, contacting the DNA polymerase with a template, a primer that binds to the template, and a collection of nucleotides including at least one acyclonucleotide, and incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides. The claims require that the utilized DNA polymerase be a member of the relevant class of DNA polymerases by specifying both a level of overall sequence identity to a member of the class and the presence of the correlated motif. Specifically, the claims specify that the DNA polymerase as an amino acid sequence that shows at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4, and further includes a 15 amino-acid motif that is identical to one of SEQ ID NOs 5-22 except that it contains up to three (i.e., 0-3) amino acid substitutions as compared with the SEQ ID NO.

The claimed methods are described, inter alia, in original claims 9, 10; page 19, lines 19-20; page 31, lines 22-28; page 32, lines 1-3; and Table 3 on pages 20-21 of the specification. Support for claim 32 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; page 19, lines 19-20; page 31, lines 22-28; page 32, lines 1-3 and lines 10-16; and Table 3 on pages 20-21. Support for claim 33 is found in the specification as originally filed, inter alia, in original claim 10 and at page 19, lines 18-20. Support for claim 34 found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 35 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 36 is found in the specification as originally filed, inter alia, in

original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 37 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 38 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 39 is found in the specification as originally filed, inter alia, in original claim 19. Support for claim 40 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 41 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 42 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 43 is found in the specification as originally filed, inter alia, in original claims 13 and 18.

**GROUND OF REJECTION TO BE REVIEWED ON APPEAL**

The grounds of rejection to be reviewed on appeal are:

- (1) Are claims 32-42 invalid for lack of written description under 35 U.S.C. § 112?
- (2) Are claims 32 -42 invalid for lack of enablement under 35 U.S.C. § 112?



## **ARGUMENT**

Claims 32 and 39 stand or fall together and each of claims 33, 34, 35, 36, 37, 38, 40, 41, and 42 stands or falls alone for grounds of rejection (1) and (2) to be reviewed upon appeal, as indicated below.

### **Ground of Rejection 1:**

*Claims 32 and 39 are not invalid for lack of written description*

Pending claims 32-42 stand rejected for lack of written description. The Examiner states that claims 32-42 contain subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventors, at the time the application was filed, had possession of the claimed invention. This rejection is respectfully traversed. Reconsideration and withdrawal is requested.

The written description requirement serves both to satisfy the inventor's obligation to disclose the technologic knowledge upon which the patent is based, and to demonstrate that the patentee was in possession of the invention that is claimed. *Capon v. Eshhar*, 418 F.3d 1349, 1357 (Fed. Cir. 2005). To satisfy the written description requirement, the applicant does not have to utilize any particular form of disclosure to describe the subject matter claimed, but the description must clearly allow persons of ordinary skill in the art to recognize that he or she invented what is claimed. *Carnegie Mellon Univ. v. Hoffmann La Roche Inc.*, 541 F.3d 1115, 1122 (Fed. Cir. 2008) (quoting *In re Alton*, 76 F.3d 1168 (Fed. Cir. 1996)). In other words, the applicant must 'convey with reasonably clarity to those skilled in the art that, as of the filing date sought, he or she was in possession of the invention,' and demonstrate that by disclosure in the specification of the patent. *Id.* Such disclosure need not recite the claimed invention *in haec verba*, but it must do more than merely disclose that which would render the claimed invention obvious. *Univ. of Rochester v. G.D. Searle & Co.*, 358 F.3d 916, 923 (Fed. Cir. 2004). The descriptive text needed to meet the written description requirement varies with the nature and scope of the invention at issue, and with the scientific and technologic knowledge already in existence. *Capon*, 418 F.3d at 1357.

The present claims recite methods comprising steps of providing a particular type of DNA polymerase, contacting the DNA polymerase with a template, a primer that binds to the template, and a collection of nucleotides including at least one acyclonucleotide; and incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides. Claim 32 specifies that the DNA polymerase has an amino acid sequence that shows at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4, and further includes a 15 amino acid motif that is identical to one of SEQ ID NOs 5-22 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO. The recited 15 amino acid motifs are shown in Table 3 of the specification at pages 20-21.

The Examiner has maintained that the written description requirement is not met for the scope of DNA polymerases encompassed by the claims. Appellants explain below that a structure/function relationship has been established for the DNA polymerases recited in the claimed methods, and that written description for the claims is more than satisfied under *Invitrogen Corp. v. Clontech Labs, Inc.*, 77 USPQ2d 1161 (Fed. Cir. 2005), and under the U.S. Patent and Trademark Office Written Description Training Materials (Revision 1, March 25, 2008).

*A Structure/Function Relationship has been established.*

The Examiner maintains the rejection for lack of written description on the ground that “applicants have not related the subgenus of structure to the acyclonucleotide incorporation function” (Office Action mailed April 21, 2009, page 4). Appellants respectfully disagree with this assertion. The disclosure of the specification, working examples, and declaratory evidence demonstrates a relationship between the structure recited in the claims and acyclonucleotide incorporation function. The claims require that the DNA polymerase have an amino acid sequence with at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4 (Vent™). The claims also require that the DNA polymerase include a 15 amino acid motif that is identical to one of SEQ ID NOs 5-22, or has up to three amino acid substitutions.

The specification explains that proteins can display sequence similarity over short stretches of primary amino acid sequence (specification, page 14). These patches are thought to occur most often at essential protein interfaces, such as those involved in catalysis, substrate binding, or protein-protein recognition. The degree of sequence similarity, particularly in conserved sequence motifs, is predictive of the degree to which the proteins will behave similarly in both physical properties and catalytic function (specification, page 14, lines 10-16). The claims include just such a motif, by requiring that the DNA polymerase include a 15 amino acid motif that is identical to one of SEQ ID NOs 5-22, or has up to three amino acid substitutions. The sequences of the 15 amino acid motifs and the DNA polymerase in which each is found are shown in Table 3 of the specification at pages 20-21. Each motif is within a conserved region having a role in substrate binding, known as “motif B” as defined by Delarue et al. (*Protein Eng.* 3:461-467, 1990; see citation to Delarue et al. in the specification at page 21 under Table 3; Delarue et al. was submitted with the Information Disclosure Statement filed on May 9, 2002, and is attached as **Exhibit A**). Delarue et al. does not recognize or discuss acyclonucleotide activity of any DNA polymerases. However, Delarue et al. indeed indicates that motif B is involved in DNA polymerase function. In the Discussion section, Delarue et al. states:

***From structure to function. Considerable biochemical evidence points to the importance of [motifs A, B and C] in the DNA polymerase activity. A synthesized E. coli pol I oligopeptide corresponding to the N-terminal-most two-thirds of the loop region connecting helices O and P (motif B-see Figure 4) has been shown to bind deoxynucleotide triphosphate substrates of pol I as well as duplex DNA (Mildvan, 1989)”*** (Delarue et al., page 465, right col., lines 9-15; emphasis added).

This structure/function relationship between motif B and polymerase activity is further confirmed in declaratory evidence submitted during prosecution of the present application. In the Declaration by Dr. William Jack, filed on May 4, 2006 (“the Jack Declaration”; copy attached as **Exhibit B**), it states that Dr. Jack and colleagues have published articles in peer reviewed journals discussing the physical basis for the preferential incorporation of acyclonucleotides and enhanced incorporation with Vent A488L and 9°N 485L DNA polymerase mutants, citing to Gardner et al. (*J. Biol. Chem.* 279(12):11834-11842, 2004; Gardner et al. was submitted with the Jack Declaration and is attached as **Exhibit B**). Gardner et

al. shows an alignment of Family B DNA polymerases in Figure 1. As is clear from the Figure, the “Region III” active site overlaps with the 15 amino acid motif recited in Appellants’ claims. As provided in the Jack Declaration, Gardner discusses the physical basis for incorporation of acyclonucleotides at page 11841. This discussion mentions the A288 residue in Vent™, which is in the active site and in the 15 amino acid motif in Appellants’ claims. A relationship between “Region III”, containing the 15 amino acid motif, and polymerase function, had been previously noted, e.g., in Hopfner et al. (*Proc. Nat. Acad. Sci. USA* 96:3600-3605, 1999; Hopfner et al. was submitted with the Jack Declaration and is attached as **Exhibit B**). Hopfner et al. reports the crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. Hopfner et al. provide a structure based sequence alignment of archaeal family B polymerases, and show that Region III (which contains the 15 amino acid motif) is in the active site of these enzymes (see Hopfner et al., page 3603, col. 1, Figure 3, and col. 2, third full paragraph). Hopfner et al. discusses the conserved KX<sub>3</sub>NSXYGX<sub>2</sub>G motif, which is a sub-motif within Appellants’ claimed 15 amino acid motif, in the section entitled “Polymerase Active Site”, noting that it and a second motif “form the bottom of the nucleotide-binding site” (Hopfner et al., page 3603, right col., lines 31-32). The subgenus of structure (i.e., the 15 amino acid motif) is clearly related to function.

Although there was recognition in the art that conserved motifs found in polymerases are involved in polymerase activity, it is Appellants who recognized and now claim a method of using a specific genus of polymerases which possess acyclonucleotide incorporation function. The set of 15 amino acid motifs specified by SEQ ID NOs 5-22 and recited in the claims are highly related to each other. SEQ ID Nos 6-17 differ from SEQ ID No 5 by three or fewer residues. SEQ ID Nos 18 and 20-22 differ from SEQ ID Nos 5 by six or fewer residues. Motifs of polymerases having sequences sharing less than 30% overall identity with Vent™ (and thus which are outside the scope of the claims) have motifs which differ from SEQ ID No 5 by seven or more residues (see, e.g., SEQ ID Nos 23-30 at page 21, Table 3 of the specification).

A structure/function relationship is not only supported by an understanding of the 15 amino acid motif and its role in enzymatic function. It is also supported by Appellants’ working examples. Every DNA polymerase tested that meets the structural requirements of the claims has acyclonucleotide incorporation activity. Indeed, four different DNA polymerases,

Vent™, Deep Vent™, *Pfu*, and 9N™, showed the ability to incorporate acyclonucleotides (specification, Example 6). Two variants of these enzymes, Vent™/A488L, and 9N™/A485L, were also shown in incorporate acyclonucleotides (specification, Example 11). By contrast, Thermosequenase, which is a Taq DNA polymerase variant that lacks the 15 amino acid motif required by the claims, showed a much stronger preference for dideoxyoligonucleotides over acyclonucleotides (specification, Examples 5 and 12). The application therefore establishes the correlation between the sequence motif and the function recited in the claims.

In addition, the Jack Declaration includes an Appendix with data confirming that an archaeon Family B polymerase from *Methanococcus maripaludis*, having 41% sequence identity with Vent DNA polymerase, utilizes acyclonucleotides as a substrate (Jack Declaration Appendix I, attached as **Exhibit B**). Thus, support for a relationship between the DNA polymerases recited in the claims and acyclonucleotide incorporation function has been provided in by information in the specification regarding sequence similarity and function, exemplification of a relationship between the claimed structure and function in the specification, and data and information provided with the Jack Declaration.

The Examiner maintains his rejection without offering any reason *why* the claimed structure/function relationship allegedly has not been established. For example, in the Office Action mailed May 29, 2008, the Examiner said that “[w]hile Applicants comments regarding the homogeneity shared between this group of polymerases continues to be acknowledged, such is acknowledged in light of the degree of the vast majority of DNA polymerases, many of which have a high degree of homogeneity and not all of which share the ability to incorporate acyclonucleotides into a DNA fragment” (Office Action mailed May 29, 2008, page 4). Appellants have related specific structural features (overall sequence identity and the presence of a 15 amino acid motif in the active site of the enzyme) to function (acyclonucleotide incorporation function). The Examiner has provide no *reason* to doubt Appellants correlation. The Examiner is not entitled to substitute his personal skepticism for statements and evidence provided by the Appellants.

*Written description support for the claims is met under Invitrogen Corp. v. Clontech Labs, Inc. 77 USPQ2d 1161 (Fed. Cir. 2005).*

Relevant legal precedent also confirms that the written description requirement is satisfied for the present claims in view of the present specification. The decision in *Invitrogen Corp. v. Clontech Labs, Inc.* 77 USPQ2d 1161 (Fed. Cir. 2005) requires a finding that the claims are adequately described. To emphasize this point, Appellants reiterate a close comparison between *Invitrogen* and the present claims here. The claim at issue in *Invitrogen* read:

1. An isolated polypeptide having DNA polymerase activity and substantially reduced RNase H activity, wherein said polypeptide is encoded by a modified reverse transcriptase nucleotide sequence that encodes a modified amino acid sequence resulting in said polypeptide having substantially reduced RNase H activity, and wherein said nucleotide sequence is derived from an organism selected from the groups consisting of a retrovirus, yeast, *Neurospora*, *Drosophila*, primates and rodents.

The specification supporting the claim had only a single example of a polymerase having the recited activity. The court found that the claim met the written description requirement because, (1) at the time of the invention, sequences of reverse transcriptase (RT) genes were known; (2) members of the RT gene family shared significant homologies from one species to another; (3) the written description taught that the invention can be applied to RT genes of other retroviruses; and (4) the specification cited references providing the known nucleotide sequences of those genes.

It must be noted that, unlike the claim in *Invitrogen* which recites no structural limitations, the pending claims include explicit recitation of structural features (overall homology and a 15 amino acid motif). The present specification provides six specific examples of DNA polymerases that fall within the claims. As for the other factors from *Invitrogen*, (1) sequences of many DNA polymerases were known when the present application was filed; and (2) members of the DNA polymerase gene family share significant homologies from one species to another. See the present specification, e.g., at page 3, lines 8-21; and page 10, line 12, to page 15, line 34. For (3), the written description of the present case clearly teaches that the invention can be applied to DNA polymerases other than the ones specifically exemplified. See, for example, page 19, lines 15-27, which teaches:

The similarity of incorporation patterns with these selected enzymes suggests that not only these archaeon DNA polymerases, but a larger family of DNA polymerases could share the ability to incorporate acyclo to a greater extent than dideoxy terminators. Since *Pfu*, Deep Vent® and 9°N™ DNA polymerases have greater than about 70% sequence identity with Vent DNA polymerase, other enzymes with equivalent or greater identity can reasonably be expected to perform as Vent® (exo-) DNA polymerase in this invention. Notably, those enzymes for which no significant sequence similarity is found (i.e., Family A DNA polymerases such as Taq) do not perform in similar ways in the current invention. This fact leads us to believe that archaeon enzymes showing intermediate identity, namely those between about 20 and 70% identity are reasonably candidates for this invention.

As to (4), the specification cites references providing the known sequences of such other DNA polymerases (see, for example, page 10, line 22; page 14, line 18; page 14, line 19; page 15, lines 19-24). Moreover, the sequences of other DNA polymerases are known and need not be fully presented in the specification to satisfy the written description requirement. See *Capon*, 418 F.3d at 1358.

Appellants maintain that, with regard to every relevant fact relied upon by the court, the present case has at least as much, or more description than was provided in *Invitrogen*.

The Examiner disputes this point because the claims encompass incorporation of acyclonucleotides into DNA and

[t]his is not a property of a DNA polymerase that is well known in the art and the applicants have not adequately described this supposedly new function of a specific sub-genus of DNA polymerases. This is in contrast to the claims of *Invitrogen* in which the homologies of the encompassed DNA polymerases were high and that region responsible for reduced RNase H activity in each of these DNA polymerases known such that the encompassed DNA polymerase variants known. (Office Action mailed April 21, 2009, page 5).

Appellants explain in detail the relationship between structure provided and acyclonucleotide function above. As discussed, Appellants have demonstrated (through several examples) that DNA polymerases that do have the claimed sequence do have the recited activities, and a DNA polymerase that does not have the claimed sequence does not have the recited activity.

Moreover, the fact that the present claims recite use to perform a newly discovered function (incorporation of acyclonucleotides) does not distinguish the present case from *Invitrogen*, as asserted by the Examiner. The claims in *Invitrogen* also related to DNA polymerases that have a new function (reduced RNase H activity). The Examiner is correct that the *region* of DNA polymerase sequence that was responsible for RNase H activity was previously known. As discussed above, the relevant region of DNA polymerases (region III) involved in the present claims was also known (and known to be important for activity, just not for this activity). The present specification demonstrates that this known region is important for a new activity, much like the specification in *Invitrogen* demonstrated that changes in a known region could reduce activity. Closer factual scenerios in fact would be difficult to find!

Furthermore, Appellants fail to see how acyclonucleotide function of the DNA polymerases renders this case distinguishable from *Invitrogen*. In that case, a single example of an enzyme having a desired function (reduced RNase H activity) was adequate to support the claims.

Appellants' recognition of a class of polymerases which incorporate acyclonucleotides is new, and Appellants have linked the functional activity with structure and a characterized structural, functional motif (i.e., the 15 amino acid motif). There is no basis for distinguishing the present case from *Invitrogen*. The Examiner suggests that *Invitrogen* is not applicable because "the homologies of the encompassed DNA polymerases were high." Yet the *Invitrogen* claim is completely devoid of structural limitations, and recites polymerases from organisms as diverse as viruses, yeasts, and primates! If unspecified sequences from such varied species have "high" homology in the Examiner's view, Appellants fail to understand how homologies between sequences encompassed by the present claims, which recite concrete structural limitations, are not also "high."

In a further attempt to distinguish *Invitrogen*, the Examiner stated that

the description held by *Invitrogen* is specific to the claims of *invitrogen* [sic], based upon the specification and art as well as a. Actual reduction to practice, b. Disclosure of drawings or structural chemical formulas, c. Sufficient relevant identifying characteristics, such as: Complete structure, ii. Partial structure, iii. Physical and/or chemical properties, iv. Functional characteristics when coupled with a known or disclosed correlation between function and structure, d. Method



of making the claimed invention, e. Level of skill and knowledge in the art and f. Predictability in the art. (Office Action mailed April 21, 2009, carryover paragraph from pages 5-6).

Legal decisions would be meaningless as precedent if they could be applied only to a single set of facts. Appellants have provided a close comparison of (i) the facts in the *Invitrogen* case and the (stronger) facts here; and (ii) the claims of the present application and a claim from *Invitrogen* for which written description was affirmed. There has been no showing that *Invitrogen*'s claimed genus all had "high" homology or "known" function such that the present claims can be distinguished from the case. No other bases for finding *Invitrogen* inapplicable have been offered.

*Written description support for the claims is met under the U.S. Patent and Trademark Office Written Description Training Materials*

The Examiner compared the present claims to the U.S. Patent and Trademark Office Written Description Training Materials (hereinafter, the "Guidelines") and found lack of description in Appellants' claims compared to claim 2 in Example 11 of the Guidelines because "claim 2 is drawn to a nucleic acid having 85% identity to a specific sequence, a partial structure. This is relative to the instant claims which require even less partial structure of 30% identity." (Office Action mailed April 21, 2009, page 7).

The Examiner has not analyzed Appellants' claims in view of the knowledge of DNA polymerase structure and the requirement of a conserved motif which is associated with enzymatic function. Claim 2 in Example 11 of the Guidelines concerns a claim to nucleic acid encoding hypothetical polypeptide having "activity X". In contrast to the present claims, the hypothetical polypeptide encoded by the nucleic acid does not share significant sequence identity with any known polypeptide or polypeptide family. Also unlike the present claims, the specification for this hypothetical example discloses only a single nucleic acid sequence that encodes a polypeptide having "activity X". Any comparison of the present claims to Example 11 should take these facts into consideration. Another important factor for analysis in Example 11

is the presence of a disclosed or art-recognized correlation between structure and function. Appellants have provided this correlation.

Example 5 of the Guidelines presents a fact pattern much more analogous to Appellants' claims, and is a more appropriate basis for comparison. Example 5 concerns a claim to an "isolated protein comprising Protein A," wherein Protein A includes the amino acid sequence of SEQ ID NO:1, has the ability to bind and activate Protein X, and is purified by a recited set of conditions. The sequence of SEQ ID NO:1 in this hypothetical claim has 10 amino acids. Likewise, Appellants' claims recite DNA polymerases that include a 15 amino acid motif and have a specific binding and activity function, which is the ability to incorporate acyclonucleotides in a polymerase extension reaction. The polymerases are not defined by purification conditions. However, significant structural definition for the polymerases is provided by requiring at least 30% identity to SEQ ID NO:4.

In the hypothetical fact pattern set forth for Example 5, claim 1, the specification fails to disclose the complete structure of Protein A and it fails to disclose any art recognized correlation between the structure of the claimed protein and its function of binding and activating Protein X. Nonetheless, written description is affirmed for the claim because the specification discloses a partial (10 amino acid) sequence of Protein A and because relevant identifying characteristics are provided in the form of its ability to bind and activate Protein X, and purification features.

If anything, the present specification provides more description support for the claims than is provided for claim 1 of Example 5 of the Guidelines. Appellants' specification describes examples of complete structures for polymerases that fall within the claims. Appellants' 15 amino acid motif imposes greater structural definition for a polymerase than the 10 amino acid sequence defining the hypothetical polypeptide of Example 5. Appellants' polymerases possess a binding ability and activity (acyclonucleotide incorporation) which is just as well defined as those of the hypothetical polypeptide of Example 5. Whereas no correlation of protein structure with function is provided in Example 5, Appellants' provide detailed structure/function correlation, as set forth above. In this aspect, Appellants provide more support than the Guidelines require. Another factor favoring support for the hypothetical polypeptide was the specification's disclosure of methods for isolating the polypeptide and a working example showing the polypeptide was successfully isolated. Appellants' have also shown that one of skill

in the art can make and use polypeptides as claimed, and that polypeptides have the recited function.

*Claim 33 is not invalid for lack of written description*

Claim 33 stands rejected for lack of written description. Claim 33 specifies that the DNA polymerase has an amino acid sequence that shows at least 70% overall identity with that of SEQ ID NO:4. Because this claim requires a higher overall identity to SEQ ID NO:4, the genus of polymerases encompassed by the claim is smaller than that of claim 32. Thus, the level of description required is reduced as compared with claim 32. Appellants' specification demonstrates that multiple polymerases within the genus possess acyclonucleotide function. (Appellants emphasize that polymerases from the broader genus have this function as well; see the Jack Declaration, Appendix I, which shows that a *Methanococcus* DNA polymerase having only 41% sequence identity to Vent™ DNA polymerase incorporates acyclonucleotides more efficiently than dideoxynucleotides.) Even if claim 32 were not fully supported by the specification (which Appellants do not concede), claim 33 would be.

*Claim 34 is not invalid for lack of written description*

Claim 34 stands rejected for lack of written description. Claim 34 specifies that the 15 amino acid motif is identical to one of SEQ ID Nos 5-22. Given the further limitation on the sequence of the motif (i.e., such that the motif does not include amino acid substitutions), the genus of polymerases encompassed by the claim is smaller than that of claim 33. The level of description required for this claim is reduced as compared with claim 32. Even if claim 32 were not fully supported by the specification, claim 34 would be.

*Claim 35 is not invalid for lack of written description*

Claim 35 stands rejected for lack of written description. Claim 35 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 15-17, except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO. Because it covers fewer motifs, this

claim refers to a genus of polymerases that is smaller than that encompassed by claim 32. The level of description required to support this claim is less than required for claim 32.

*Claim 36 is not invalid for lack of written description*

Claim 36 stands rejected for lack of written description. Claim 36 specifies that the 15 amino acid motif is identical to one of SEQ ID Nos 5-17. The genus of polymerases encompassed by this claim is even smaller than that of claim 32 and requires less description to be adequately supported.

*Claim 37 is not invalid for lack of written description*

Claim 37 stands rejected for lack of written description. Claim 37 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 5-8 except that it may contain up to three amino acid substitutions. Again, the genus of polymerases encompassed by this claim is even smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully supported by the specification.

*Claim 38 is not invalid for lack of written description*

Claim 38 stands rejected for lack of written description. Claim 38 specifies that the amino acid motif is identical to one of SEQ ID NOs 5-8. The genus of polymerases encompassed by this claim is smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully supported by the specification.

*Claim 40 is not invalid for lack of written description*

Claim 40 stands rejected for lack of written description. Claim 40 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-22. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully supported by the specification.

*Claim 41 is not invalid for lack of written description*

Claim 41 stands rejected for lack of written description. Claim 41 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-17. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully supported by the specification.

*Claim 42 is not invalid for lack of written description*

Claim 42 stands rejected for lack of written description. Claim 42 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID Nos 5-8. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully supported by the specification.

In conclusion, the provided teachings in the specification, examples, sequences, declaratory evidence, and data are more than sufficient to describe function and support description of the claims. The Examiner has not established otherwise. For reasons set forth above, withdrawal of the rejection of claims 32 and 39 as allegedly lacking written description is respectfully requested.

## **Ground of Rejection 2:**

*Claims 32 and 39 are not invalid for lack of enablement*

Pending claims 32 and 39 stand rejected for lack of enablement. The Examiner states that the specification, while being enabling for a method comprising providing a DNA polymerase selected from the group consisting of Vent™, Deep Vent™, *Pfu*, and 9°™ or the specifically disclosed variants of claim 43, “does not reasonably provide enablement for any method comprising providing a DNA Polymerase having an amino acid sequence that shows a mere 30% overall identity with that of SEQ ID NO:4 and further includes a 15 amino-acid motif that is identical to SEQ ID NO:5 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO...” (Office Action mailed April 21, 2009, pages 8-9). The Examiner stated that “determination of those DNA polymerases having the desired biological characteristics is unpredictable and the experimentation left to those skilled in the art is unnecessarily, and improperly, extensive and undue” (Final Office Action mailed April 21, 2009, pages 11). Appellants have previously reviewed the factors set forth in *In re Wands* (858 F.2d 731, 8 USPQ2d 1400, Fed. Cir. 1988) with respect to the present claims and review them here, in response to the Examiner’s assertion that Appellants’ burden has not been met.

First, Appellants address the Examiner’s comments regarding *Wands* factor (2), Amount of Direction or Guidance. In the Final Office Action mailed April 21, 2009, the Examiner maintained that guidance was lacking as to DNA polymerases which have the ability to incorporate acyclonucleotides into a DNA template, and requested clarification as to how the 15 amino acid motif correlates with acyclonucleotide function (Office Action mailed April 21, 2009, page 12). As explained above in the arguments for written description, the 15 amino acid motif is a highly conserved motif in the active site of family B DNA polymerases which plays a role in substrate binding. The Examiner disputes a structure/function correlation because “applicants have not disclosed such a single motif but rather continue to refer to any of a number of motifs or variants thereof” (Office Action mailed April 21, 2009, page 12). Some variability within the genus of motifs is permitted, given that variable polymerases share acyclonucleotide incorporation function. For example, both 9°N polymerase and Vent™ incorporate

acyclonucleotides, although their 15 amino acid motifs differ by three amino acids (compare SEQ ID NO 5 and SEQ ID NO 7 at page 20, Table 3 of the specification). A *Methanococcus maripaludis* DNA polymerase having a more divergent sequence also possesses acyclonucleotide incorporation activity. The claims do require a degree of conservation of sequence, which is clearly expressed in the claims. The fact that variable polymerases share a specific function does not render them “unpredictable.”

As to the (1) Quantity of Experimentation Necessary, and (3) Presence or Absence of Working Examples, Appellants reiterate that one of ordinary skill could make and test all polypeptides within the scope of the claims to determine their ability to extend a DNA primer or incorporate acyclonucleotides (including to determine their ability to preferentially select acyclonucleotides). Appellants’ working examples include demonstration of activity of multiple species of DNA polymerases set forth in the specification and in declaratory evidence discussed herein.

As to (5) State of the Prior Art, and (7) Predictability of the Art, Appellants note, and the Examiner has acknowledged, that the prior art with regard to DNA polymerases and their classification is extensive. However, Appellants disagree with the Examiner’s assertion that “determination of those DNA polymerases having the desired biological characteristics is unpredictable and the experimentation left to those skilled in the art is unnecessarily, and improperly, extensive and undue” (Office Action mailed April 21, 2009, page 11). Appellants have identified polymerases which have acyclonucleotide incorporation function by virtue of structural and physical characteristics distinctive of well-characterized DNA polymerases. These characteristics include overall sequence identity to a polymerase, and the presence of a conserved motif. Appellants have shown that all members tested within the genus of polymerases have the recited activity. The Examiner’s only discernible reason for declaring these features unpredictable is the breadth of the genus of polymerases. This improperly disregards Appellants’ demonstration of activity for multiple species. It also disregards the fit of Appellants’ observed activity with well characterized classification schemes for DNA polymerases (among which one finds substantial variability despite conserved nucleotide polymerase activity).

As to the (4) Nature of the Invention and (8) Breadth of the Claims, Appellants reiterate that DNA extension reactions are well within the skill of those of ordinary skill. As part of their invention, Appellants have described a class of DNA polymerases that can incorporate acyclonucleotides, and have shown function for six different species within the class. Given the demonstrated correlation of structure with function and other reasons provided above, Appellants disagree with the Examiner's assertions that the scope of the claims is not enabled.

As to (6) Relative Skill of those in the Art, Appellants submit, and the Examiner has agreed, that the relative skill of those in the art is very high.

*Claim 33 is not invalid for lack of enablement*

Claim 33 stands rejected for lack of enablement. Claim 33 depends from claim 32 and specifies that the DNA polymerase has an amino acid sequence that shows at least 70% overall identity with that of SEQ ID NO:4. Because this claim requires a higher overall identity to SEQ ID NO:4, the breadth of the claim is smaller than that of claim 32. The scope of enablement provided by the disclosure is more than sufficient to support the scope of this claim, not least because multiple polymerases that fall within the claimed genus are exemplified.

*Claim 34 is not invalid for lack of enablement*

Claim 34 stands rejected for lack of enablement. Claim 34 depends from claim 32 or 33 and specifies that the 15 amino acid motif is identical to one of SEQ ID Nos 5-22. This further limitation on the sequence of the motif (i.e., such that the motif does not include amino acid substitutions) provides a claim of smaller breadth than claim 32 and which is more than supported by the disclosure.

*Claim 35 is not invalid for lack of enablement*

Claim 35 stands rejected for lack of enablement. Claim 35 depends from claim 32 or 33 and specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 15-17, except that it



contains up to 3 amino acid substitutions as compared with the SEQ ID NO. This claim covers fewer motifs than claim 32 and is enabled for its full scope.

*Claim 36 is not invalid for lack of enablement*

Claim 36 stands rejected for lack of enablement. Claim 36 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 5-17. Again, the genus of polymerases encompassed by this claim is even smaller than that of claim 32 and is enabled by the disclosure provided.

*Claim 37 is not invalid for lack of enablement*

Claim 37 stands rejected for lack of enablement. Claim 37 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 5-8 except that it may contain up to three amino acid substitutions. The genus of polymerases encompassed by this claim is even smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully enabled by the specification.

*Claim 38 is not invalid for lack of enablement*

Claim 38 stands rejected for lack of enablement. Claim 38 specifies that the amino acid motif is identical to one of SEQ ID NOs 5-8. The genus of polymerases encompassed by this claim is smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully enabled by the specification.

*Claim 40 is not invalid for lack of enablement*

Claim 40 stands rejected for lack of enablement. Claim 40 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-22. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully enabled by the specification.

*Claim 41 is not invalid for lack of enablement*

Claim 41 stands rejected for lack of enablement. Claim 41 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-17. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully enabled by the specification.

*Claim 42 is not invalid for lack of enablement*

Claim 42 stands rejected for lack of enablement. Claim 42 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID Nos 5-8. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is enabled for its full scope.

In light of the above, Appellants submit that claims 32 -42 satisfy the enablement requirement. Allowance of the claims is requested.

Date: February 2, 2010

PATENT GROUP  
CHOATE, HALL & STEWART LLP  
Two International Place  
Boston, MA 02110  
Tel: (617) 248-5000  
Fax: (617) 502-5002  
patentdocket@choate.com

Respectfully submitted,

/Margo H. Furman/  
Margo H. Furman, PhD, JD  
Registration Number 59,812  
Tel. No. (617) 248-4073

## **CLAIMS APPENDIX**

1-31. (Canceled)

32. (Previously presented) A method comprising steps of:

providing a DNA polymerase having an amino acid sequence that shows at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4, and further includes a 15 amino-acid motif that is identical to one of SEQ ID NOs 5-22 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO;

contacting the DNA polymerase with a template, a primer that binds to the template, and a collection of nucleotides including at least one acyclonucleotide; and

incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides.

33. (Previously presented) The method of claim 32, wherein the DNA polymerase has an amino acid sequence that shows at least 70% overall identity with that of SEQ ID NO:4.

34. (Previously presented) The method of claim 32 or claim 33, wherein the 15 amino-acid motif is identical to one of SEQ ID NOs 5-22.

35. (Previously presented) The method of claim 32 or claim 33, wherein the 15 amino-acid motif is identical to one of SEQ ID NOs 5-17 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO.

36. (Previously presented) The method of claim 35, wherein the 15 amino acid motif is identical to one of SEQ ID NOs 5-17.

37. (Previously presented) The method of claim 32 or 33, wherein the 15 amino acid motif is identical to one of SEQ ID NOs 5-8 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO.
38. (Previously presented) The method of claim 37, wherein the 15 amino acid motif is identical to one of SEQ ID NOs 5-8.
39. (Previously presented) The method of claim 32 or 33, wherein the step of incubating comprises incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides, and preferentially incorporates acyclonucleotides.
40. (Previously presented) The method of claim 32 or 33, wherein the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-22.
41. (Previously presented) The method of claim 35, wherein the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-17.
42. (Previously presented) The method of claim 37, wherein the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-8.
43. (Previously presented) The method of claim 32 or 33 wherein the DNA polymerase is Vent<sup>TM</sup>, Deep Vent<sup>TM</sup>, 9<sup>o</sup>N, *Pfu*, Vent<sup>TM</sup>/488L, or 9<sup>o</sup>N/485L.

## **EVIDENCE APPENDIX**

Appellants had provided the following evidence during prosecution of the instant application:

**Exhibit A:** Delarue et al., *Protein Eng.* 3:461-467, 1990. This reference was cited in the Information Disclosure Statement and Form PTO-1449 filed on May 9, 2002, and was entered into the record on May 13, 2002. The Form PTO-1449 was initialed by the Examiner on September 29, 2004, confirming that the reference was entered into the record.

Delarue et al. is attached hereto at pages 31-37.

**Exhibit B:** Declaration of William Jack, accompanying references and Appendix I. The Declaration was submitted with four references, listed below, and Appendix I along with a response to Office Action filed May 4, 2006, and was entered into the record in PAIR on May 9, 2006 as the entry designated "Rule 130, 131 or 132 Affidavits." Entrance into the record was confirmed by the Examiner's reference to this Declaration on page 3 of the Advisory Action mailed on July 5, 2006.

The Declaration of William Jack is attached hereto at pages 38-45.

Rodriguez et al., *J. Mol. Biol.* 299:447-462, 2000, is attached hereto at pages 46-61.

Gardner et al., *J. Biol. Chem.* 279(12): 11834-11842, 2004, is attached hereto at pages 62-70.

Hashimoto et al., *J. Mol. Biol.* 306:469-477, 2001, is attached hereto at pages 71-79.

Zhao et al., *Structure* 7(10):1189-1199, 1999, is attached hereto at pages 80-90.

Hopfner et al., *Proc. Nat. Acad. Sci. USA* 96:3600-3605, 1999, is attached hereto at pages 91-96.

Appendix I is attached hereto at pages 97-99.

**RELATED PROCEEDINGS APPENDIX**

Not applicable.

## EXHIBIT A

Protein Engineering vol.3 no.6 pp 461-467, 1989

### An attempt to unify the structure of polymerases

Marc Delarue, Olivier Poch<sup>1</sup>, Noel Tordo<sup>2</sup>, Denis Moras and Patrick Argon<sup>3</sup>

<sup>1</sup>Department of Crystallography and Biophysics, IMAC to CNRS, 15 rue des Sciences, 91800 Breteuil, France; <sup>2</sup>Centre de Recherches, Institut Pasteur, 25 rue de Valenciennes, 92204 Paris Cedex 15, France and <sup>3</sup>Biological Molecular Biology Laboratory, Institut 19.2205, Menlo Park, CA 94025, USA

With the great availability of sequences from RNA- and DNA-dependent RNA and DNA polymerases, it has become possible to delineate a few highly conserved regions for various polymerase types. In this work a DNA polymerase sequence from bacteriophage SPQ2 was found to be homologous to the polymerase domain of the Klenow fragment of polymerase I from *Escherichia coli*, which is known to be closely related to those from *Staphylococcus pneumoniae*, *Thermus aquaticus* and bacteriophages T7 and T3. The alignment of the SPQ2 polymerase with the other five sequences considerably improved the conserved motifs in these proteins. Three of the motifs matched reasonably all the conserved motifs of another DNA polymerase type, characterized by human polymerase  $\alpha$ . It is also possible to find these three motifs in monomeric DNA-dependent RNA polymerases and two of them in DNA polymerase  $\beta$  and DNA terminal transferases. These latter two motifs also matched two of the four motifs recently identified in RNA-dependent polymerases. From the known tertiary architecture of the Klenow fragment of *E. coli* pol I, a spatial arrangement can be implied for these motifs. In addition, numerous biochemical experiments suggesting a role for the motifs in a common function (NTP binding) also support these inferences. This speculative hypothesis, attempting to unify polymerase structure at least locally, if not globally, under the pol I fold, should provide a useful model to direct mutagenesis experiments to probe template and substrate specificity in polymerases.

**Key words:** catalytic domain/DNA polymerase/RNA polymerase/sequence/structure

#### Introduction

The number of available protein sequences is growing rapidly due to the facility of nucleotide sequencing techniques. One protein class often studied and sequenced is the polymerase family which is central to the duplication and expression of genes. Polymerases can use RNA or DNA as a template (RNA- or DNA-dependent); the products can also be RNA or DNA. Polymerases are found both in eukaryotes and prokaryotes, though sequencing efforts have often concentrated on those from viruses. One way to use the information contained in all these sequences is to try to align them and thereby allocate them amongst related families and subfamilies. This has been achieved for DNA-dependent DNA polymerases, where three main subfamilies have been identified. One of them contains the Klenow fragment of *Escherichia coli* polymerase I, whose three-dimensional structure is known (Ollis *et al.*, 1985a), and

polymerases from phages T7 (Ollis *et al.*, 1985b; Argon *et al.*, 1986) and T3 (Leavitt and Ro, 1989), and from *Thermus aquaticus* (Lapeze *et al.*, 1989) and *Staphylococcus pneumoniae* (Lapeze *et al.*, 1989). This family will be referred to as the pol I family. For another set of DNA-dependent DNA polymerases (Wang *et al.*, 1989), as those homologous to the human polymerase  $\alpha$  (hereafter referred to as pol  $\alpha$ ), more than 10 sequences from various species are known. A third subfamily of DNA-dependent DNA polymerase, hereafter called the pol  $\beta$  type, has only two members: DNA polymerase  $\beta$  (Matsukage *et al.*, 1987) and terminal transferase (Peterson *et al.*, 1981; Norwood, 1986; Zaslavskii *et al.*, 1986). Until now, just one DNA-dependent DNA polymerase sequence, from the SPQ2 bacteriophage (Thelen and Rutberg, 1984; Jung *et al.*, 1987), resisted alignment with any of the three aforementioned types.

Clearly, the aim of these alignments, apart from evolutionary implications, is the identification of the regions essential for polymerase function, since these sequence segments should appear as the most conserved. Generally, a great number of aligned sequences will ensure sufficient variability to identify the functionally required regions. For the pol I type, the five previously aligned (Ollis *et al.*, 1985b; Argon *et al.*, 1986; Leavitt and Ro, 1989; Lapeze *et al.*, 1989) sequences are sufficiently close to us to allow confident delineation of the absolutely required motifs.

In the present work, it has been found that the polymerase from bacteriophage SPQ2 can be aligned, using a sensitive method, with the polymerase portion of the Klenow fragment in the C-terminal part of the protein [the N-terminal domain has a 3'-5' exonuclease function (Thelen *et al.*, 1989)]. The total alignment of the C-terminal part of the six proteins of the pol I type is presented. The relationship of SPQ2 polymerase with those from phages T7 and T3, *Thermus aquaticus*, *T. aquaticus* and *E. coli* is sufficiently distant that highly conserved regions can now be reduced to five in number. Interestingly, three of the five regions match reasonably with the three most conserved motifs of DNA-dependent DNA pol  $\alpha$ , suggesting that the two polymerase types may share a common tertiary fold, or at least contain similar local tertiary architecture required for similar functions. These motifs are likely to represent modules required for the polymerase structure and activity.

Searches were then performed to detect such sequence patterns in other polymerase families. All these motifs could be found in DNA-dependent RNA polymerases that consist of only one subunit (see Moras *et al.*, 1987); two motifs were found in pol  $\beta$ s, in the same linear arrangement and maintaining the strictly conserved residues. In addition to this, an examination of several aligned RNA-dependent RNA polymerases as well as reverse transcriptases, for which four highly conserved motifs have been highlighted (Krause and Argon, 1984; Poch *et al.*, 1989), allows the suggestion that two of these motifs could match the two motifs shared by DNA pol  $\beta$ s, pol  $\alpha$ s and pol Is. These sequence similarities are further supported by a statistically significant alignment between the entire polymerase domains of two members of these different families, namely a DNA-dependent DNA

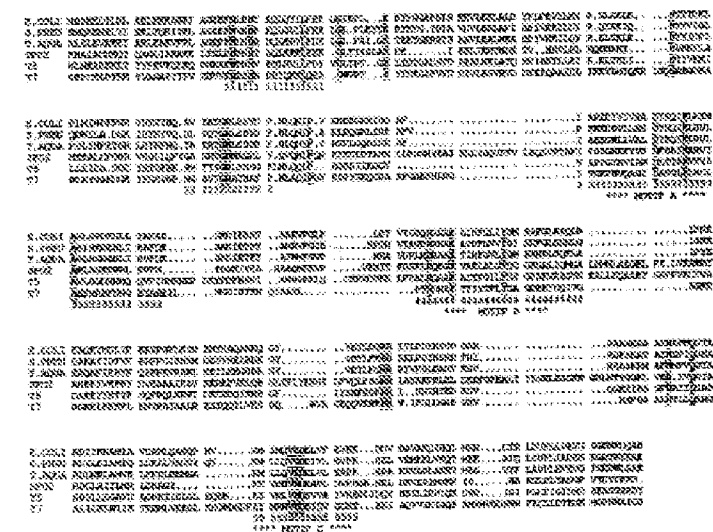
© Oxford University Press

461

polymerase and an RNA-dependent DNA polymerase. The oligonucleoside of the C-terminal most motif, comprising two acidic groups, is in agreement with those suggested by Argon (1982) amongst RNA-dependent RNA polymerases, reverse transcriptase and RNA-dependent DNA pol. *in situ*.

## Reading

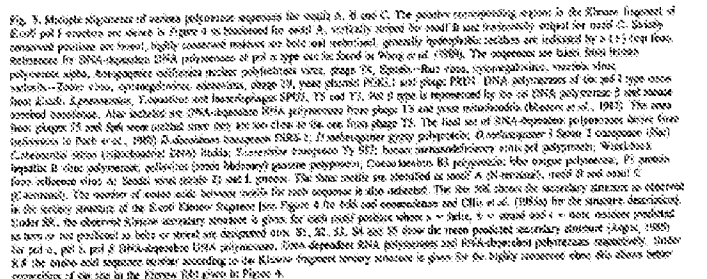
Pairwise sequence comparisons were done by a search procedure based on residue characteristics (Argers, 1987). The resulting alignment of the E-terminal polypeptide portion of the Eukaryotic fragment of DNA-dependent DNA polymerase  $\beta$  from *Leish* and polypeptides from baculoviruses 73, T3, *S. pomonensis* and *T. leucopis* are similar to the ones given by Argers et al. (1986b), Loefer and his (1989), Lopez et al. (1989) and Lopez et al. (1989) respectively. The 50% polypeptide sequence from baculoviruses 50552 and also compared to the E. coli, *S. pomonensis*, T3 and 73 polypeptides. The search matrix and



422



alignment was obtained by manual adjustment of the different test classes pairwise alignments. A conservation profile resulting from this alignment was also calculated (data not shown); this profile is based on a three-residue window and the score is simply the normalized sum of the matrix elements corresponding to the residues observed in all the different pairwise alignments. The





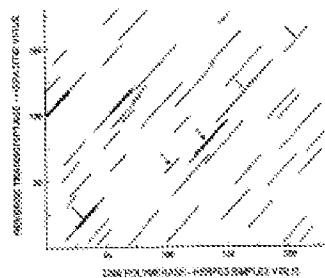


Fig. 3. Homology search (Argos, 1987) between Wadsworth sequence B RNA dependent DNA polymerase and P-loop virus DNA dependent DNA polymerase. The search window length ranged from 6 to 28 in step of 2; the search pools are plotted over the entire window length, with the larger value connecting with overlap occurs from the nullspace. The peak value (5) was scored as a number of identical positions for those the score was by each window length. The first column: 3.5 to 5.5; the second: 4.0 to 5.5; the third: 4.5 to 5.5; the fourth: 5.0 to 5.5; the fifth: 5.5 to 5.5. The plot used as a basis to effect the alignment shown in Figure 6 is indicated by arrows.

common. The distances between these motifs are quite variable. However, within each polymerase type, there is already great variation in the distances between conserved regions. Moreover, it would seem unlikely that these of five most conserved regions in the polymerase domain of pol  $\beta$  (Figure 2) should be aligned with the three most conserved regions of DNA pol as easily by chance. A different piece of evidence pointing to a possible link between the pol I and the pol  $\alpha$  types comes from sequence similarities within the exonuclease domain of the TS polymerase and pol  $\alpha$  (Suzuki and Ito, 1989). It is then reasonable to expect that polymerase domains to be related after all, this observation has been recently extended to all exonuclease domains of pol  $\alpha$  and pol  $\beta$ , including SPOB (Bernard *et al.*, 1989). In addition, the secondary structure predictions, especially for motifs A and C, point to supersecondary structures for pol  $\alpha$  that coincide with those observed in the pol I Klenow structure.

*From the sequence to structure.* The three motifs A, B and C are clustered and consecutive in the C-terminal fraction of the Klenow fragment and correspond to structural features likely to interact with the DNA (Ols *et al.*, 1983a). Figure 4 highlights the portions of the Klenow pol I tertiary structure corresponding to the three shared motifs. Motif A, characterized by a conserved Asp and Ser, corresponds to strand 9 and helix 1. Motif B, with conserved Lys, Tyr and Gly encompasses most of helix 2 and the following long loop region. Motif C delineates strands 12 and 13 with the mostly conserved negative charges contained in the connecting loop of the  $\beta$ -hairpin. Since the N-terminal Asp of the doublet is the only one universally conserved, it is presumed to be catalytically more important. It is clear that the structural segments A and B are likely to come into contact with DNA and that the strand 12–13 hairpin loop could place in contact one catalytic residue in the polymerase active site. Model building of DNA into the Klenow fragment structure also supports the importance of these regions (Ols *et al.*, 1983a,b; Warwicker *et al.*, 1985). Furthermore, the C<sub>6</sub> atom of motif C Asp82 is

within 3.5 Å of that from Asp702, conserved in motif A (C<sub>6</sub> atom coordinates given in Brookhaven database (Bernstein *et al.*, 1977), file 1D8PG). Their spatial proximity would allow both to participate in catalysis. Region 2 of pol  $\beta$  (sequence TCR between strands 7 and 8—see Figure 2 and 4) is also located in the vicinity of this area but no sequence homology with the pol  $\alpha$  type could be found for this region. Region 1 falls in an undefined part of the electron density map (see Ols *et al.*, 1983a; Figure 4). *From structure to function.* Certain biochemical evidence points to the importance of these three motifs in the DNA polymerase activity. A synthesized 21-residue pol I oligopeptide corresponding to the N-terminal-most two-thirds of the loop region connecting helices 1 and 2 (motif B—see Figure 4) has been shown to bind deoxynucleotide triphosphate substrates of pol I as well as duplex DNA (Mokawa, 1986). Furthermore, photo-affinity labeling with 8-azido-dATP identifies Tyr766 as a residue in the active site (Joyce and Joyce, 1987) which, also part of this motif, accounts to chemical labeling, this time using pyridyl phosphate (Beno and Modak, 1987). A factor synthesized peptide (Shinkagawa *et al.*, 1988; McDevitt, 1989) corresponding to helix 1 and strands 12 and 13 (motif C) was not found to bind the pol I substrate, although it apparently retains its proper folding; however, this result does not exclude that this peptide, although unable to bind dNTP on its own, can co-operate with other regions of the native protein and the put part of the binding site. In fact, His861 of pol I, which is in the loop joining strands 12 and 13 and sequentially adjacent to the conserved Asp882 of motif C, has been shown to be involved in the binding of [<sup>32</sup>P]dATP (Pascaly *et al.*, 1987). In addition, the beginning of helix 3, which is close in space to the end of helix 2 (motif B) and motifs A and C, can also be labeled by phenylglyoxal (Mokawa *et al.*, 1988).

Similarly, for the pol  $\alpha$  type, protein studies of the Herpes simplex virus polymerase (Larder *et al.*, 1987a) revealed that four out of the six identical drug-resistant polymerase subunits cluster in motifs A and B. The drug used was a dNTP analog. Other mutants involved in drug and substrate recognition were also mapped by Gibbs *et al.* (1982) in motifs A and B. This is in agreement with a central role of these motifs in dNTP binding for the pol  $\alpha$  type, as is the case of the pol  $\beta$  type.

Finally, for the third DNA polymerase type (pol  $\gamma$ ), a chemical affinity labeled [<sup>32</sup>P]8-azido-dATP peptide in the terminal exonuclease allowed the mapping of the dNTP binding site; this peptide contained the sequence GRHNDQ that is part of motif C (Kjawa *et al.*, 1989) and that resembles both YGHDD and VGHDF. This supports the alignment given in Figure 1, even though motif B could not be found with certainty in this pol  $\gamma$  type of DNA polymerase. Furthermore, the number of residues between motifs A and C in the structure is 101 residues with the corresponding number in the Klenow pol I averaging about 120 residues.

*Structural implications of the distance variability between motifs.* Figure 5 shows the number of residues contained between each of the motifs for all the DNA pol  $\alpha$  sequences and for the Klenow polymerase 1 domain. Between motifs A and B a comparable number of residues is found for both pol  $\alpha$  and  $\beta$  types, while between B and C the pol  $\alpha$  sequences generally contain considerably fewer amino acids than their positive Klenow counterparts. This region of the structure encompasses helix 3, strands 10 and 11, and helix 2. Apart from the beginning of helix 2, these regions are unlikely to be in contact with the DNA, according to model building studies (Warwicker *et al.*, 1985; Figure 4), suggesting that this region could be considerably

[illegible]

Fig. 6. Alignment of sequence 8 (reverse, nonconsensus (200)-dependent DNA polymerase from *Shigella flexneri* (Bergdoll)) with the T60-dependent DNA sequence from *Shigella flexneri* (Bergdoll). (%), nucleic identity; (//), no mutation in the HpaII sequence, namely, in single base pair mutations (TTCG→GTCG; GCG→GCG; GCG→GCG).

aligned with one another approximately horizontally. The shortest segment between nucleotides B and C is 25 atoms wide in the horizontal plane or sequence; a peptide of this size would severely hinder the  $\pi$ - $\pi$  interaction of the C<sub>6</sub> systems of the flanking nucleotides B and Arg193 at the 3' terminus of model C, i.e., the nucleoside (see following table) of the C<sub>6</sub> system of nucleotides B and Arg193. In fact, given the 40 Å distance between the C<sub>6</sub> atoms of Arg193 and Arg193 in the C<sub>6</sub> model of tertiary structure (Jöres et al., 1985b), 19 methylene in helical conformation (mean C-C distance of 1.5 Å) and eight in a coil structure (mean C-C distance of 2.9 Å) could open the required 45 Å. This "helical model" would allow the maintenance of the 12-membered ring of helix Q, whose sequence is also necessarily conserved in the six type I DNA polymerase sequences (see Figure 2).

The distance between motifs A and B can also vary from 50 to 120 residues in the put re family; this region corresponds to helices M and N of the Kinase structure. In the put 1 family, this distance ranges from 20 to 60 residues. An inspection of Figure 4 suggests the possibility of insertions in this region.

Dissolution of *Asiaticus*

In this family, three conserved regions that included the three motifs A, B and C could be found. Invariant residues are maintained and the residues highly between these conserved sequences with those of DNA-dependent DNA polymerases. However, since this family is composed of five members with only two sufficiently distant ones, more sequences are needed in order to narrow the number of conserved regions that are truly functionally important.

Alcázar-dependent polymers:

Of the four motifs of viral RNA-dependent RNA polymerases, as well as reverse-transcriptase sequences recently identified by Pock et al. (1989), two motifs A and C are RNA-dependent polymerases. These motifs are the only sequence features shared by all these sequences. The 51% identity alignment of the hepatitis B reverse transcriptase from Woodchuck (Gelbert et al., 1982) with the Hepatitis simplex virus DNA pol. (Gilbert et al., 1985) strongly supports the relationship between RNA- and RNA-dependent polymerases. The phylogeny of these two sequences is also sensible as evolutionary genes, because reverse transcriptases have already been postulated to be an essential intermediary step to go from an RNA-dependent RNA polymerase to a DNA-dependent DNA polymerase (Lassus et al., 1989). Furthermore, it should be pointed out that hepatitis B viruses are the only viruses encoding reverse transcriptase activity that have a DNA genome.

A size-directed stereoregular experiment has pointed to the catalytic importance of the Asp-Arg doublet in itself C of RNA-dependent polymerases (Suzuki and Hirotsu, 1987). In addition, site-specific mutagenesis of the amino terminus

deficiency virus reverse transcriptase revealed that mutation of the first conserved Asp of motif C, one that completely destroyed activity (Larder et al., 1987b). The mutation of the other strictly conserved Asp residue, in motif A, from Glu had a similar effect.

The sequence length between motifs A and C is the RNA-dependent polymerase averages  $\sim 20$  nucleotides, which is considerably shorter than the 150–120 nucleotides in the pfs, though one sequence from influenza virus contains 115 spacer residues, comparable with some pfs (Figure 3). Once again, motif B and part of the helix C and strands D and E are possible candidates for deletion. The absence of all the distances C8 residues suggest that deletions can occur in the M and N helices and in the alpha helix-like helix, U, in the family.

### Discussion and conclusion

Argon (1988) previously suggested the conservation of motif C in pol cat and RNA-dependent polymerases; we tested this observation in the p1a, pol p1a and some DNA-dependent RNA polymerases. Apparently the fully conserved form of Argon's App-Arg domain of motif C is evolutionarily more important than the second. We also confirm that there is at least another motif upstream (motif A), with another strongly conserved separate form; is a two between a beta strand and an alpha helix, in specific proximity of motif C; genetic experiments suggest that this motif might also be involved in the catalytic process. One of the many possible functional roles for these residues is that the serine conserved Arg residue in motif C, may co-operate with the use of motif A to bind a magnesium ion that would be part of the dNTP binding site. Then this would be located at the bottom of a cleft containing the DNA. This cleft, similar to the one of the Klenow fragment, is already apparent in a 4 Å. 1989. If this is true, the different families of polymerases could be the result of divergences from a common ancestor, with conserving the same folding, or the conservation of conserved residues in the same spatial arrangement in the catalytic site could be due to convergent evolution, as observed in the subtilisin and chymotrypsin active site.

In spite of wide apparent sequence variability (likely to reflect a very fast net divergence), it is very tempting to suggest that the conserved, structurally important gene products sharing fast matrix polymerases may hold the key to the higher-order architecture of the *E. coli* pil. Consistent with this hypothesis is the experimental observation that it is possible to change the topology or substrate specificity of conserved polymerase types *III*  $\beta$ g<sup>+</sup> (replaced by  $\beta$ g<sup>-</sup>) (see Lazarini et al., 1988). This model, however, cannot be viewed as speculative, even though several lines of evidence point to its underlying consistency. We believe it deserves attention, because site-directed mutagenesis experiments aimed at probing the metabolic site and topology specificity should benefit from our considerations, which give a possible structural framework for

Received on November 24, 1999; accepted on February 28, 2000

## 467

- Received on November 24, 1999; accepted on February 28, 2000

## **EXHIBIT B**

Docket No.: NEB-165-PJ5

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS: Jack et al. EXAMINER: Hutson

SERIAL NO.: 10/089,027 ART UNIT: 1652

DATE FILED: March 26, 2002

TITLE: Incorporation of Modified Nucleotides By Archaeon DNA Polymerases  
And Related Methods

Mail Stop AF  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

---

### **DECLARATION UNDER 37 C.F.R. §1.131**

As a below named inventor, I hereby declare that:

1. My name is Dr. William Jack, Research Director for the DNA Enzymes Division at New England Biolabs Inc. My resume is attached.
2. I have been studying the structure and function of DNA polymerases for over 16 years.
3. I was a member of the group of scientists at New England Biolabs that isolated, characterized, and cloned the first hyperthermophilic archaeal DNA polymerase. Our continuing work with archaeon DNA polymerases identified a surprisingly homogeneous set of enzymes. We claimed this group of DNA polymerases in US Patent 5,500,363. In this patent, the United States Patent and Trademark Office recognized the validity of our claim to a class of archaeon DNA polymerases defined by the DNA encoding the enzyme and its

**BEST AVAILABLE COPY**

ability to hybridize under defined conditions to various specified DNA sequences. The group was exemplified by T.litoralis (Vent), GBD (Deep Vent), and 9°N DNA Polymerases.

4. We also found that this group of polymerases had a high degree of amino acid sequence identity. A comparative three-dimensional alignment of members of this group of enzymes showed a high degree of structural conservation, consistent with the observed high degree of primary amino acid sequence identity/similarity. See for example, Vent (Rodriguez, et al., 2000), Tgo (Hopfner, et al., 1999), D. Tok (Zhao, et al., 1999), and KOD (Hashimoto, et al., 2001) DNA Polymerases.

5. The structural equivalence of this group of polymerases is further supported by experiments reported in Example 10 of the above application in which we show that mutation of an analogous residue in Vent and 9°N DNA Polymerases yields enzymes with equivalent acyclo nucleotide incorporation efficiencies.

6. We discovered that this group of enzymes is capable of efficiently utilizing acyclo nucleotides as substrates. We demonstrated this property using four examples of polymerases within this tightly defined group. Any molecular biologist of ordinary skill in the art would expect from these findings that this property would occur in all members of the enzyme group defined above.

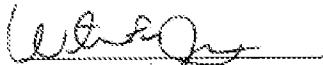
7. Additionally, my colleagues and I have published articles in peer reviewed journals discussing the physical basis for the preferential incorporation of acyclo nucleotides, and also for the enhanced incorporation with Vent A488L and 9°N A485L DNA Polymerase mutants. See Gardner, et al. (2004) on page 11841, column 1, paragraph 2 and page 11841, column 2, paragraph

1, respectively.

8. I assert that the combination of the high degree of homogeneity in DNA and amino acid sequences of archaeon DNA polymerases, plus the structural evidence that modification of specific amino acids alters enzyme specificity, would be sufficient to assure a person of ordinary skill in the art that the class of polymerases as defined above will interact with acyclonucleotide substrates as shown in the above application.

9. To further support the above statements, we have conducted additional experiments to confirm that archaeon Family B polymerases with an amino acid sequence identity of greater than 30% can utilize acyclonucleotides as a substrate. This data is attached to the present declaration as appendix 1.

9. I further declare under penalty of perjury pursuant to laws of the United States of America that the foregoing is true and correct and that the Declaration was executed by me on:

  
William E. Jack

Date: 4 May 2006



References:

- A. Gardner, C.M. Joyce, W.E. Jack (2004) *J. Biol. Chem.* 279, 11834-11842.
- H. Hashimoto, M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, T. Inoue, Y. Kai (2001) *J. Mol. Biol.* 306, 469-477.
- K.P. Hopfner, A. Eichinger, R.A. Engh, F. Laue, W. Ankenbauer, R. Huber, B. Angerer (1999) *Proc. Natl. Acad. Sci. USA* 96, 3600-3605.
- A.C. Rodriguez, R.W. Park, D. Mao, L.S. Beese (2000) *J. Mol. Biol.* 299, 447-462.
- Y. Zhao, D. Jeruzalmi, I. Moarefi, L. Leighton, R. Lasken, J. Kuriyan (1999) *Structure* 7, 1189-1199.

**William Eric Jack**

New England Biolabs  
240 County Road  
Ipswich, MA 01938  
(978) 386-7237  
(978) 921-1350 (fax)  
email: jack@neb.com

---

**RESEARCH INTERESTS**

Enzymatic and structural aspects of protein-nucleic acid interactions. Thermostable DNA polymerase kinetics and function.

**RESEARCH EXPERIENCE**

**New England Biolabs (Beverly, MA).**

2005-present Division Head, DNA Enzymes

1987-present Senior Staff Scientist

Research: Kinetic characterization of thermostable DNA polymerases.  
Creation and characterization of DNA polymerase variants with altered substrate recognition. Over-expression and characterization of restriction and modification enzymes.

2000-present New England Biolabs Institutional Biosafety Committee Chair

**Rockefeller University (NY, NY)** Laboratory of Biochemistry and Molecular Biology.

1983-1987 Postdoctoral Fellow in the laboratory of R.G. Roeder.

Research: Structural and functional characterization of wild type and mutant forms of *Xenopus* RNA polymerase III transcription factor A. Glucocorticoid hormone-induced transcription enhancement *in vitro*.

**Duke University (Durham, NC)** Department of Biochemistry.

1977-1983 Graduate Student in the laboratory of P. Modrich.

Research: Kinetics and thermodynamics of DNA site location, recognition and cleavage by *EcoRI* endonuclease.

**EDUCATION**

Doctor of Philosophy (Biochemistry), Duke University, 1983 (Paul Modrich, advisor).

Bachelor of Arts (Chemistry), *Magna Cum Laude*, University of Utah, 1977.

**TRAINING**

2000

Sixth National Symposium on Biosafety: Prudent Practices for the New Millennium (Conducted by the Centers for Disease Control and Prevention)

## **PUBLICATIONS**

- "Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase," Andrew F. Gardner and **William E. Jack**, *J. Biol. Chem.* **279**, 11834-11842 (2004).
- "Acylic and Dideoxy Terminator Preferences Denote Divergent Sugar Recognition by Archaeon and Tsg DNA Polymerases," Andrew F. Gardner and **William E. Jack**, *Nucleic Acids Res.*, **30**, 605-613 (2002).
- "The Kinetic Mechanism of EcoRI Endonuclease," David J. Wright, **William E. Jack** and Paul Modrich, *J. Biol. Chem.* **274**, 31896-31902 (1999).
- "Determinants of Nucleotide Sugar Recognition in an Archaeon DNA Polymerase," Andrew F. Gardner and **William E. Jack**, *Nucleic Acids Res.* **27**, 2545-2553 (1999).
- "Photochemically Initiated Protein Splicing", Sandra N. Cook, **William E. Jack**, Xiaofeng Xiong, Lora E. Dunley, Jonathan A. Ellman, Peter G. Schultz, and Christopher J. Noren, *Angew. Chem. Int. Ed. Engl.* **34**, 1629-1636 (1995).
- "Three-dimensional Structure of the Adenine-specific DNA Methyltransferase M-TaqI in Complex with the Cofactor S-adenosylmethionine", J. Latahn, J. Granzin, G. Schluckebier, D.P. Robinson, **W.E. Jack**, I. Schildkraut, and W. Saenger, *Proc. Natl. Acad. Sci. USA* **91**, 10957-10961 (1994).
- "Expression, Purification, and Crystallization of Restriction Endonuclease PvuII With DNA Containing Its Recognition Site", K. Balendiran, Joseph Borventre, Roger Knott, **William Jack**, Jack Benner, Ira Schildkraut, and John E. Anderson, *Proteins: Structure, Function, and Genetics* **19**, 77-79 (1994).
- "Protein Splicing Elements: Introns and Exons-A Definition of Terms and Recommended Nomenclature", Francine B. Perler, Elaine O. Davis, Gary E. Dean, Frederick S. Gimble, **William E. Jack**, Norma Neff, Christopher J. Noren, Jeremy Thirner and Marlene Belfort, *Nucleic Acids Res.* **22**, 1125-1127 (1994).
- "Characterization of a DNA Polymerase from the Hyperthermophile Archaea *Thermococcus lituralis*", Hunmin Kong, Rebecca B. Kucera and **William E. Jack**, *J. Biol. Chem.*, **268**, 1965-1975 (1993).
- "Protein Splicing Removes Intervening Sequences in an Archaea DNA Polymerase", Robert A. Hodges, Francine B. Perler, Christopher J. Noren and **William E. Jack**, *Nucleic Acids Res.* **20**, 6153-6157 (1992).
- "Intervening Sequences in an Archaea DNA Polymerase Gene", Francine B. Perler, Donald G. Comb, **William E. Jack**, Laurie S. Merus, Bojin Qiang, Rebecca B. Kucera, Jack Benner, Barton E. Slafko, Donald O. Nwankwo, S. Kay Hempstead, Clotilde K.S. Carlow and Holger Jannasch, *Proc. Natl. Acad. Sci. USA* **89**, 5577-5581 (1992).
- "Overexpression, Purification and Crystallization of BsuHI Endonuclease", **William E. Jack**, Lucia Greenough, Lydia F. Dorner, Shuang-yong Xu, Teresa Strelacka, Ansel K. Aggarwal and Ira Schildkraut, *Nucl. Acids Res.* **19**, 1825-1829 (1991).

"Nucleotide Sequence of the *FokI* Restriction-Modification System: Separate Strand-Specificity Domains in the Methyltransferase", Mary C. Looney, Laurie S. Moran, **William E. Jack**, George R. Fechery, Jack S. Benner, Barton E. Slafko and Geoffrey G. Wilson, *Gene* 80, 193-208 (1989).

"M.FokI methylates adenine in both strands of its asymmetric recognition sequence", David Landry, Mary C. Looney, George R. Fechery, Barton E. Slafko, **William E. Jack**, Ira Schildkraut, and Geoffrey G. Wilson, *Gene* 77, 1-10.

"Mechanism of specific site location and DNA cleavage by *EcoRI* endonuclease", Brian J. Terry, **William E. Jack** and Paul Modrich, *Gene Amplif Anal.* volume 5, pp. 103-118 (1987).

"Facilitated diffusion during catalysis by *EcoRI* endonuclease. Nonspecific interactions in *EcoRI* catalysis", Brian J. Terry, **William E. Jack** and Paul Modrich, *J Biol. Chem.* 260: 13130-13137 (1985).

*Participation of Outside DNA sequence in the EcoRI Endonuclease Reaction Pathway*, **William E. Jack**, dissertation (Duke University, Durham, NC) (1983).

"Thermodynamic parameters governing interaction of *EcoRI* endonuclease with specific and nonspecific DNA sequences", Brian J. Terry, **William E. Jack**, Robert A. Rubin and Paul Modrich, *J Biol. Chem.*, 258: 9826-9825 (1983).

"Involvement of outside DNA sequences in the major kinetic path by which *EcoRI* endonuclease locates and leaves its recognition sequence", **William E. Jack**, Brian J. Terry, and Paul Modrich, *Proc Natl. Acad Sci U S A* 79: 4010-4014 (1982).

"DNA determinants important in sequence recognition by *EcoRI* endonuclease", A-Lien Lu, **William E. Jack** and Paul Modrich, *J Biol. Chem.* 256: 13200-13206 (1981).

"Structures and Mechanisms of *EcoRI* DNA Restriction and Modification Enzymes", **William E. Jack**, Robert A. Rubin, Andrea Newman and Paul Modrich, *Gene Amplif. Anal.* volume 1, p 165-179 (1981).

#### **US PATENTS**

"Use of site-specific nicking endonucleases to create single-stranded regions and applications thereof," **William E. Jack**, Ira Schildkraut, Julie Forney Menin, US Patent 6,660,473, December 9, 2003.

"Recombinant thermostable DNA polymerase from archaeobacteria," Donald G. Corb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,834,285, November 10, 1998.

"Modified proteins comprising controllable intervening protein sequences or their elements methods of producing same and methods for purification of a target protein comprised by a modified protein," Donald G. Corb, Francine B. Perler, **William E. Jack**, Ming-Qun Xu, Robert A. Hodges, Christopher J. Noren, Shaorong S.C. Cheng, Eric Adam, Maurice Southworth, US Patent 5,834,287, November 10, 1998.

- "Recombinant thermostable DNA polymerase from archaeobacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,500,363, March 19, 1996.
- "Modification of protein by use of a controllable intervening protein sequence," Donald G. Comb, Francine B. Perler, **William E. Jack**, Ming-Qun Xu, Robert A. Hodges, US Patent 5,496,714, March 5, 1996.
- "Recombinant thermostable DNA polymerase from archaeobacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,352,778, October 4, 1994.
- "Purified thermostable DNA polymerase obtainable from thermococcus litoralis," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,322,785, June 21, 1994.

**JMB**



## Crystal Structure of a Pol $\alpha$ Family DNA Polymerase from the Hyperthermophilic Archaeon *Thermococcus* sp. 9°N-7

A. Chapin Rodriguez†, Hee-Won Park‡, Chen Mao and Lorena S. Beese\*

Department of Biochemistry  
Duke University Medical  
Center, Durham  
NC 27710, USA

The 2.35 Å resolution crystal structure of a pol  $\alpha$  family (family 8) DNA polymerase from the hyperthermophilic marine archaeon *Thermococcus* sp. 9°N-7 (9°N-7 pol) provides new insight into the mechanism of pol  $\alpha$  family polymerases that include essentially all of the eukaryotic replicative and viral DNA polymerases. The structure is folded into NH<sub>2</sub>-terminal, editing 5'–3' exonuclease, and polymerase domains that are topologically similar to the two other known pol  $\alpha$  family structures (bacteriophage RB69 and the recently determined *Thermococcus gorgona*), but differ in their relative orientation and conformation.

The 9°N-7 polymerase domain structure is reminiscent of the “closed” conformation characteristic of ternary complexes of the pol I polymerase family obtained in the presence of their dNTP and DNA substrates. In the apo-9°N-7 structure, this conformation appears to be stabilized by an ion pair. Thus far, the other apo-pol  $\alpha$  structures that have been determined adopt open conformations. These results therefore suggest that the pol  $\alpha$  polymerases undergo a series of conformational transitions during the catalytic cycle similar to those proposed for the pol I family. Furthermore, comparison of the orientations of the fingers and exonuclease (pol) domains relative to the palm subdomain that contains the pol active site suggests that the exonuclease domain and the fingers subdomain of the polymerase can move as a unit and may do so as part of the catalytic cycle. This provides a possible structural explanation for the interdependence of polymerization and editing exonuclease activities unique to pol  $\alpha$  family polymerases.

We suggest that the NH<sub>2</sub>-terminal domain of 9°N-7 pol may be structurally related to an RNA-binding motif, which appears to be conserved among archaeal polymerases. The presence of such a putative RNA-binding domain suggests a mechanism for the observed autoregulation of bacteriophage T4 DNA polymerase synthesis by binding to its own mRNA. Furthermore, conservation of this domain could indicate that such regulation of pol expression may be a characteristic of archaea. Comparison of the 9°N-7 pol structure to its mesostable homolog from bacteriophage RB69 suggests that thermostability is achieved by shortening loops, forming two disulfide bridges, and increasing electrostatic interactions at subdomain interfaces.

© 2000 Academic Press

**Keywords:** Archaea; X-ray structure; replication; exonuclease; family II DNA polymerase

\*Corresponding author

†Contributed equally to the manuscript.  
Abbreviations used: pol, polymerase; Tgs, *Thermococcus gorgonae*; ddNTP, dideoxynucleotides.  
E-mail address of the corresponding author: [lsb@biochem.duke.edu](mailto:lsb@biochem.duke.edu)

### Introduction

DNA polymerases catalyze the template-directed addition of nucleotides onto the 3'-OH group of the DNA primer terminus. These enzymes replicate DNA with the required accuracy essential for geno-

0895-2688/00/030447-16 \$35.00/0

© 2000 Academic Press

non stability, but generates sufficient mutations to stimulate and maintain evolution. Unlike Eucarya and Bacteria, relatively little is known about DNA replication in Archaea (Barer *et al.*, 1996), one of the three major evolutionary lineages of life (Woese *et al.*, 1993). Archaea play a significant role in the biosphere, accounting for up to 30% of the biomass in certain Antarctic waters (De Long *et al.*, 1994), and exhibit much greater diversity than had originally been suspected (Baron *et al.*, 1996). Many characterized archaeal species are adapted to live in environments of extreme temperature, pressure, salinity, and/or pH, such as hydrothermal vents, and hot springs (Rosa & Adams, 1995).

Although archaeal cells share many morphological features with Bacteria, archaeal proteins involved in gene expression including DNA replication, transcription, and translation have been found to be similar to those from Eucarya (Edgell & Doolittle, 1997; Ishi *et al.*, 1996). In particular, most of the archaeal DNA polymerases that have been sequenced belong to the  $\alpha$ -like polymerase family (family B) that includes essentially all the eukaryotic replication and viral DNA pols (Braithwaite & Ho, 1993; Edgell *et al.*, 1997).

Crystal structures exist for DNA pols from each of four families: pol I (family A), pol II (family B), pol III (family C), and reverse transcriptase (reviewed by Joyce & Steitz, 1994; Doolittle *et al.*, 1999). Although pols from different families are structurally quite diverse, several common features have emerged. The pol domain from each resembles a right hand and may be further divided into palm, fingers, and thumb subdomains, as was originally described for the large fragment of *Escherichia coli* pol I (Klenow fragment) (Cillis *et al.*, 1985). All polymerases appear to share the same mechanism for nucleotidyl transfer involving two divalent metal ions (reviewed by Isavrigian & Steitz, 1998). In addition, based on structures containing DNA and dNTP bound to pols from pol I, pol II, and reverse transcriptase families, a conformational change in the fingers subdomain from an open to a closed conformation is proposed to occur during the catalytic cycle (reviewed by Doolittle *et al.*, 1999).

The pol  $\alpha$  family polymerases are of medical importance as targets for development of antiviral and anticancer therapeutics. For example, human pol  $\alpha$  is a target in the treatment of acute myelogenous leukemia and chronic lymphocytic leukemia (Kesting *et al.*, 1982; Robertson & Plunkett, 1995) and a variety of nucleotide analogs with antiviral activity inhibit strand elongation by pol  $\alpha$  (Huang & Plunkett, 1995; Combs & Plunkett, 1995). Furthermore, polymerases, particularly those that are thermostable, have a number of critical biotechnological applications ranging from PCR to cloning and DNA sequencing. Despite their biological, medical and biotechnological importance, the pol  $\alpha$  class of polymerases has not been structurally as well characterized as other DNA polymerase families.

Here we report the 2.25 Å resolution crystal structure of a pol  $\alpha$  family DNA polymerase from

the hyperthermophilic marine archaeon *Thermococcus* sp. 9°N-7 (9°N-7 pol). *Thermococcus* sp. 9°N-7 was isolated from a hydrothermal vent at 9° N latitude off the East Pacific Rise (Scolnick *et al.*, 1996). The structure is folded into NH<sub>2</sub>-terminal, editing 3'-5' exonuclease, and polymerase domains that are topologically similar to the two other known pol  $\alpha$  family structures (bacteriophage T4 DNA pol, actin) (Wang *et al.*, 1997) and the recently determined *Thermococcus gorgonarius* (Tgo) (Haghofer *et al.*, 1998), but differs in their relative orientation and conformation.

The pol domain structure is reminiscent of the "closed" conformation characteristic of binary complexes of the pol I polymerase family obtained in the presence of their dNTP and DNA substrates. In the apo-9°N-7 structure, this conformation appears to be stabilized by an ion pair. Thus far, the two other apo-pol  $\alpha$  structures that have been determined adopt open conformations. These results therefore suggest that the pol  $\alpha$  polymerases undergo a series of conformational transitions during the catalytic cycle similar to those proposed for the pol I family. Furthermore, comparison of the orientations of the fingers and exonuclease domains relative to the palm subdomain that contains the pol active site suggests that the exonuclease domain and the fingers subdomain of the polymerase can move as a unit, and may do so as part of the catalytic cycle. This provides a possible structural explanation for the interdependence of polymerization and editing exonuclease activities unique to pol  $\alpha$  family polymerases.

We suggest that the NH<sub>2</sub>-terminal domain of 9°N-7 pol is structurally homologous to the  $\beta$ -fold RNA-binding motif with an exposed patch of aromatic amino acid residues. Bacteriophage T4 DNA pol, actin is homologous to 9°N-7 pol, is known to bind its own mRNA and repress its own synthesis. The homology relationships to the RNA-binding motif suggest a structural basis for this regulatory mechanism. Furthermore, the conservation of this domain in other archaeal pols suggests that such outgroup regulation of pol expression may be general for archaea.

## Results and Discussion

### Crystal structure of *Thermococcus* sp. 9°N-7 pol

The structure of the full-length, 775-residue enzyme (bearing the double mutation D141A and D163A) was determined using the multiple isomorphous replacement method to a resolution of 2.25 Å. The current model has an R-factor of 23.9% ( $R_{free} = 33.8\%$ ) (Table 1). A Ramachandran plot of the model shows 86.8% of the residues in the most favored region and the remainder in additional allowed regions (12.4%) and generously allowed regions (0.8%). A total of 37 residues are not traced in the model and lie in regions of poorly defined electron density. The first of these gaps

Table 1. Chronologic data collection and refinement strategy

[illegible]

48



occurs at the bottom of the palm domain (residues 868-875), and the remainder are within the thumb region that is frequently observed to be partially disordered in x-ray polymerase structures, as is also the case here (e.g., Ollis *et al.*, 1985; Kiefer *et al.*, 1997). Although no disulfide bridges were included in the refinement, four Cys residues showed anomalous peaks in a difference Fourier map and side-chain distances and angles consistent with two disulfide bridges (Cys428-Cys442, Cys506-Cys509).

The structure of 9°N-7 pol reveals features common to all DNA pol structures as well as those that may be unique to archaeal pols. The overall shape of the enzyme can be described as a disc with a central hole that is folded into NH<sub>2</sub>-terminal, 3'-5' exonuclease, and polymerase domains (Figure 1a) and (b). Like all other pols of known structure, the pol domain resembles a right hand and may be further divided into palm, fingers, and thumb sub-domains, as was originally described for the large fragment of *E. coli* pol I (Klenow fragment) (Ollis *et al.*, 1985). 9°N-7 pol is similar in structure to the pol  $\alpha$  family polymerase from the mesophilic bacteriophage RB69 (RB69 pol) (Wang *et al.*, 1997), although a number of these (polymorphisms) are shorter than in RB69 pol (Figure 1a). Nearly all these sequence length differences are attributable to loop segments that are fewer and shorter in the hyperthermostable 9°N-7. As was first observed in the RB69 pol structure (Wang *et al.*, 1997), the 3'-5' exonuclease domain lies on the opposite side of the palm in comparison to pol I family polymerases. This domain arrangement is also seen in 9°N-7 pol and in T4 pol (Hopfinger *et al.*, 1999), indicating that this result is likely to be general for the pol  $\alpha$  family. The structural similarity between 9°N-7 and RB69 pols is significant given the low sequence identity (20%) in all but the active-site (palm) region, where sequence identity is 42% (Figure 2). Similar results hold for sequence alignments between 9°N-7 and human pol  $\alpha$ .

#### NH<sub>2</sub>-terminal domain

Many of the members of the pol  $\alpha$  polymerase family, including archaeal pols, bacteriophage T4 and RB69 DNA pols, have an NH<sub>2</sub>-terminal domain that is not observed in the pol I family. T4 pol is known to control its synthesis in vivo by a mechanism of autogenous regulation (Tuerk *et al.*, 1990). The mRNA-binding activity has been located to within the first 105 residues of the pol (Wang *et al.*, 1996), but the structure of a fragment comprising residues 1-388 of T4 pol failed to suggest a structural basis for RNA binding (Wang *et al.*, 1996). Here, we note that certain structural similarities between the homologous region in the 9°N-7 pol and the U1A RNA-binding protein may provide a rationale for RNA binding by T4 pol.

The NH<sub>2</sub>-terminal domain of 9°N-7 pol can be considered as three modules based on compactness of folding (Figure 3(a)). The first module comprises residues 1-33, a three-stranded  $\beta$ -sheet that inter-

acts extensively with the 3'-5' exonuclease domain via predominantly electrostatic interactions. Residues 33-56 act as a flexible linker connecting the first module to the second (residues 57-123). The third module comprises residues 338-372.

The second module is folded into a  $\beta$ -sheet motif, with two short  $\beta$ -strands, 5 and 6, inserted between the second and third elements. This motif occurs in a variety of proteins, and forms the basis for the most prevalent RNA binding motif, the RNA recognition motif (RRM). The RRM is present in the RNA-binding domains of hnRNP A1, spliceosomal proteins U1A and U2B', and the sex lethal protein (Bord & Dreyfuss, 1994). Although an alignment of the NH<sub>2</sub>-terminal domains of archaeal pols (Figure 3(b)), together with T4 and RB69 pols, shows that they lack the RNP1 and RNP2 sequence motifs that characterize the RRM (Bord & Dreyfuss, 1994), a number of highly conserved and invariant residues nevertheless emerges. Most of these residues fall in a cluster on the surface of the NH<sub>2</sub>-terminal domains of 9°N-7 and RB69 pols which therefore could mark the location of an RNA binding site atop the  $\beta$ -sheet platform on the face away from helix A (Figure 3(c)).

Both a sequence alignment (Figure 3(b)) and a structural comparison (Figure 3(c)) reveal that T4 and RB69 pols lack helix A and strand 7 of the  $\beta$ -sheet motif, perhaps explaining why no suggestive structural homologues to RNA-binding folds could be identified (Wang *et al.*, 1996, 1997).

Experiments are needed to determine whether the NH<sub>2</sub>-terminal domain of 9°N-7 pol binds RNA. Although the  $\beta$ -sheet motif occurs in proteins that are not thought to interact with RNA (Bord & Dreyfuss, 1994), we find its presence in the NH<sub>2</sub>-terminal domain of 9°N-7 pol, in a region known to bind RNA in T4 pol (Wang *et al.*, 1996), to be highly suggestive of this. RNA-binding capability could hold for other archaeal pols as well, since sequence alignment of NH<sub>2</sub>-terminal domain (Figure 3(b)) suggests that they share the  $\beta$ -sheet motif.

We further speculate that just as T4 pol binds its mRNA to down-regulate its own synthesis, such autogenous regulation of pol expression might occur in archaea. Autogenous gene regulation is well documented in bacteria, and has at least one precedent in archaea. It has been identified in the synthesis of the MvaL1 ribosomal protein of *Methanococcus marisnigri* (Hansen *et al.*, 1994), and postulated for a ribosomal gene cluster from the halophile *Haloquadratum walsbyi* (Shivashankar & Drenth, 1989). It is interesting that there is no structural evidence that such regulation extends to eukaryotes, as human pol  $\alpha$  shows no significant sequence homology in the NH<sub>2</sub>-terminal sequences aligned in Figure 3(b).

#### 3'-5' Exonuclease domain

This domain is responsible for binding single-stranded DNA and excising mismatched bases in the elongated primer strand. The structure

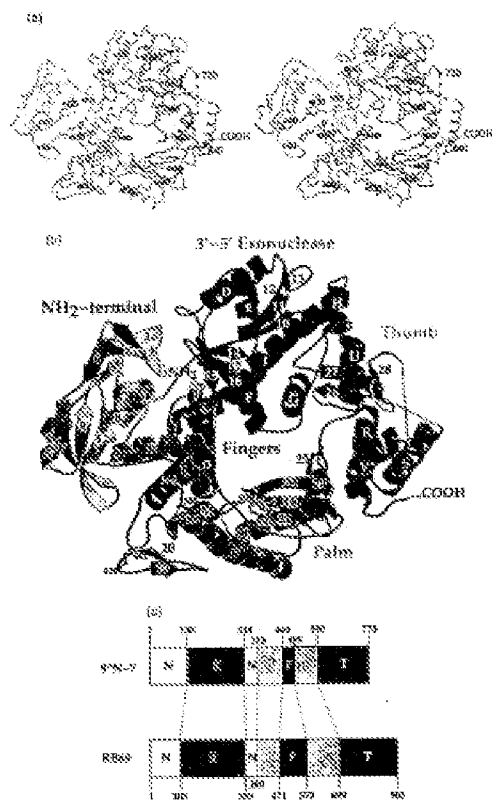


Figure 3. Structure of the *Thermococcus* sp. 97N-7 DNA polymerase. The NH<sub>2</sub>-terminal and 3'-5' exonuclease domains are colored yellow and green, respectively. The polymerase domain is divided into palm (brown), thumb (red), and finger (blue) subdomains. Three highly conserved carboxylate groups (D408, E440, D542) mark the polymerase active site. (b) Schematic of the C $\alpha$  trace. Every 40th C $\alpha$  is numbered. Broken lines indicate disordered regions of the protein. (c) Numbered diagram with secondary structure elements defined according to DSSP (Kabsch & Sander, 1983). 16H<sub>2</sub>-terminal domain: 1, 1-10; 2, 18-23; 3, 23-31; 4, 37-43; A, 48-61; 5, 55-58; 6, 61-64; 7, 67-75; 8, 78-86; 9, 92-101; 10, 106-118; C, 118-123; J, 147-148; K, 159-163; 3'-5' exonuclease domain: 10, 127-144; 11, 157-163; 12, 168-173; 13, 181-183; 14, 187-191; 15, 195-208; E, 215-225; 16, 240-244; 16, 247-251; 17, 256-260; 18, 260-266; G, 275-283; H, 292-300; I, 305-307. Polymerase domain: L, 374-376; 18, 397-404; M, 408-414; 19, 431-433; 20, 440-442; N, 448-456; O, 475-494; P, 507-512; 21, 515-516; 22, 543-547; 12, 553-557; 23, 578-584; 24, 583-598; 25, 603-606; R, 610-613; A, 636-651 (649-651 disordered); T, 660-660 (disordered); 26, 682-685; U, 677-688; 27, 689-703; 28, 714-716; V, 721-734; W, 742-745. (d) Schematic representation of the full-length of *Thermococcus* sp. 97N-7 700kDa bacteriophage RB69 DNA polymerase. The domain boundaries for 97N-7 pol were determined based upon a structure-based sequence alignment with RB69 pol (Figure 7) as defined for the RB69 pol (Yang *et al.*, 1997).

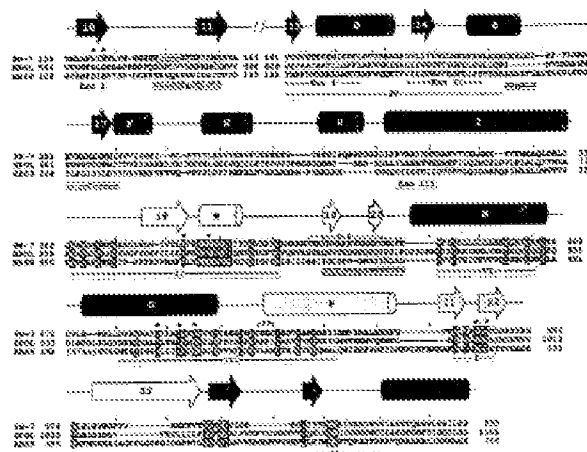


Figure 2. A three-way partial sequence alignment of *Thermococcus* sp. 97N-7 pol (874-7), RB69 pol (8869), and human pol  $\alpha$  (8406). Dashes indicate gaps in the alignment, and segments not aligned are represented as amino acid residues spaced within brackets. Ticks mark every 50 spaces. The 97N-7 and RB69 pol alignment is based upon the crystal structure. The RB69 and 8869 alignment is from Wang *et al.* (1997), except for a few short segments assigned based upon the three sequences shown here. Indicated below the sequences and boxed in yellow are consensus motifs in the exonuclease (Blanco *et al.*, 1992) and polymerase (Wang *et al.*, 1998) domains. The secondary structure elements in 97N-7 pol as defined by DSSP are given above the sequences. The structural elements are colored according to the scheme described in legend in Figure 1. Shown in purple in the 97N-7 pol sequence are the archaeal polymerase motifs described by Edgell *et al.* (1997). Residues within the polymerase domain that are invariant in the three sequences plus those residues discussed in the section on dNTP binding, blue asterisks. The two disulfide bridges in the palm (C439/C447, C766/C780) are shown schematically.

reported here is that of a mutant of 97N-7 pol lacking detectable exonuclease activity which was engineered to prevent degradation of DNA substrates during subsequent recrystallization experiments. This "97N-7mut" pol was obtained by making two point mutations (D141A, E141A) in the Exo I (DxE) motif highly conserved among the 3'-5' exonuclease domains of many DNA pols (Dontyshire *et al.*, 1995; Blanco *et al.*, 1992). In the Klenow fragment (KF) of *E. coli* DNA pol I, these residues (D355, E357) are responsible for binding the catalytic metals and for hydrogen bonding with the 3'-OH of the terminal deoxynucleotide of the substrate DNA (Boese & Steitz, 1993).

Aside from loop segments that are shorter than those observed in RB69 pol (see below), the topology of the exonuclease domain in 97N-7 pol is very similar to that of RB69 pol. The domain superimposes in the central  $\beta$ -sheet, containing the active site, with a root mean square deviation (rmad) of 0.35 Å (35 C $^{\circ}$  atoms). The metal-binding residues not mutated in 97N-7mut pol, D215 and

D215, superimpose almost exactly on the corresponding RB69 pol residues (D222, D227).

It is now possible to assign a structural context to the four archaeal sequence motifs identified by Edgell *et al.* (1997). Three of the regions (A-C) lie within the exonuclease domain (Figure 2). Motif A forms part of the central  $\beta$ -sheet containing the active site; B, part of a solvent-exposed loop; and C, part of a five-stranded  $\beta$ -sheet nearly perpendicular to the central  $\beta$ -sheet. The fourth motif resides in the palm (see below).

#### Palm domain

This domain is responsible for the template-directed polymerization of dNTPs onto the growing primer strand of duplex DNA. Like other polymerases of known structure, the palm domain can be further divided into palm, fingers, and thumb subdomains. While the structure of the thumb of 97N-7 and RB69 pols are highly similar, differences exist in the palm and fingers. Some of these differ-

ences correspond to features that appear unique in archaeal pols, while others support a hypothesis that a conformational change occurs in the fingers as part of the catalytic cycle.

#### Palm subdomain

The palm, which contains the active site for polymerization, shows a high degree of structural similarity to the palm subdomain of other DNA polymerases. It is as structurally similar to pol  $\beta$  family polymerases as to those of the pol  $\alpha$  family. Its root deviation from R869 pol around the active site (blue region in Figure 4(b)) is 0.54 Å (25 C $\alpha$  atoms). Together with the Tgo pol structure (Hopfinger *et al.*, 1999), this structure confirms for archaea the conservation of a common catalytic core. A significant difference between the palm subdomains in 9<sup>HN</sup>-7 and R869 pols are the two disulfide bridges present in 9<sup>HN</sup>-7 pol, one joining Cys428 and 442 and another joining Cys436 and 439 (Figure 4(b)). Both the shortened loops and at least one disulfide bridge appear common to archaeal pols (see above). Indeed, the region containing one of the Cys residue in a disulfide bridge (C442) corresponds to the highly conserved archaeal motif D (Edgell *et al.*, 1997; Figure 2). The Tgo pol structure shows the corresponding Cys residue to be "poised" for disulfide formation, but still in reduced form.

Until recently it was believed that all pols share a catalytic "triad" of carboxylate residues in the active site in the palm (Lohman *et al.*, 1990; Wang *et al.*, 1997) since recognized that only two of the carboxylate residues are invariant. The invariant carboxylates in 9<sup>HN</sup>-7 pol are D484 and D542. The third member of the triad, present as D543 in 9<sup>HN</sup>-7 pol, is not essential; mutation of the corresponding residue (D1032N) in human pol  $\alpha$  retains catalytic function (Copeland *et al.*, 1993). D540 in 9<sup>HN</sup>-7 pol may nevertheless be involved in binding the divalent metals required for catalysis. Mg<sup>2+</sup> is normally the optimal metal for human pol  $\alpha$  activity. The pol  $\alpha$  D1032N mutant shows greater catalytic efficiency and fidelity with Mn<sup>2+</sup> rather than Mg<sup>2+</sup> (Copeland & Wang, 1993).

D540 in 9<sup>HN</sup>-7 pol interacts with the hydroxyl group of Y538 that is within hydrogen-bonding distance to D542. Substitution of this residue to Phe in human pol  $\alpha$  (Y538G) causes only minor effects on catalysis but alters the pol metal affinity akin to the pol  $\alpha$  D1032 mutation (Copeland & Wang, 1993). It seems likely that the hydroxyl moiety of Y538 in 9<sup>HN</sup>-7 pol helps to lock D540 in position for Mg<sup>2+</sup>-specific binding. Consistent with this function is the strict conservation of Y538 among pol  $\alpha$  family members (Saitoh *et al.*, 1993).

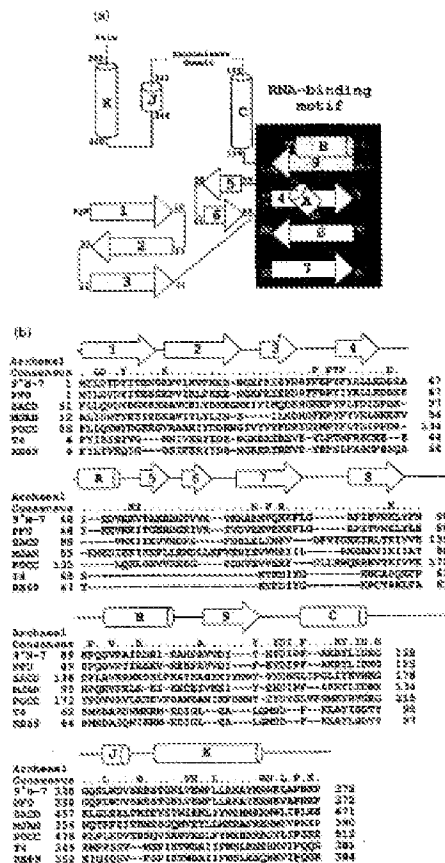
#### Fingers subdomain

The fingers subdomain of 9<sup>HN</sup>-7 differs in topology and relative conformation from R869. The fingers of 9<sup>HN</sup>-7 pol are a simple beta-strand-helix, as

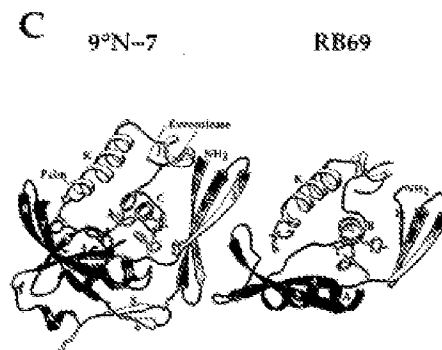
in Tgo pol (Hopfinger *et al.*, 1999), whereas in the fingers of R869 pol, the coil region is expanded with three secondary structure elements (Figures 2 and 3). The shorter fingers of 9<sup>HN</sup>-7 pol are conserved among the archaeal pols aligned by Edgell *et al.* (1997). It is possible that the fingers of archaeal pols define a minimal functional unit.

Different positions of the fingers subdomain relative to the palm are observed in the 9<sup>HN</sup>-7 and R869 pol structures (Figure 3(a)). The fingers of Tgo pol (Hopfinger *et al.*, 1999) show a position intermediate between that in 9<sup>HN</sup>-7 and R869 pols, when the palm subdomains of all three enzymes are aligned. It is interesting to note that the fingers subdomains of polymerases in the pol  $\beta$  family adopt different positions during the catalytic cycle (reviewed by Doublet *et al.*, 1999). An open position corresponds to that seen in the apoenzyme form (Silla *et al.*, 1985; Kim *et al.*, 1995; Komlev *et al.*, 1995; Kiefer *et al.*, 1997) and the form bound to duplex DNA (Birn *et al.*, 1986; Kiefer *et al.*, 1998). A closed conformation has been observed in the ternary replication complexes of bacteriophage T7 pol (Doublet *et al.*, 1998), and Klenow (Li *et al.*, 1998) with bound DNA and dNTP. An analogous conformational change has been observed in ternary complexes of human telomerase-deficient virus reverse transcriptase (Huang *et al.*, 1999) and rat pol  $\beta$  (Fellner *et al.*, 1994). In the closed conformation the fingers rotate towards the palm to form a binding pocket for dNTPs.

The differences in position of the fingers subdomain in the three pol  $\alpha$  family crystal structures suggest that the fingers of pol  $\alpha$  family pols move during catalysis, analogous to that observed for the other polymerase families. It is interesting to note that if this is the case, there must be a corresponding movement in the position of the 3'-5' exonuclease domains not required in the other polymerase families as will be discussed below. If the position of the fingers in 9<sup>HN</sup>-7 pol more closely approximates a closed conformation, it is not clear why they would adopt a position previously observed only in ternary complexes with bound dNTP and DNA. The fingers of 9<sup>HN</sup>-7 pol may be stabilized in this conformation because of a salt-bridge between E329 in the palm and K487 on helix O of the fingers. These residues are highly conserved among archaeal pols (Edgell *et al.*, 1997) and both pol  $\beta$  and pol  $\alpha$  families (Saitoh *et al.*, 1993). The corresponding salt-bridge does not form in polymerases of the pol  $\beta$  family because the fingers helix O lies too far from the palm. The fingers of Tgo pol, in fact, are rotated slightly away from the active site, relative to 9<sup>HN</sup>-7 pol, such that the E329/K487 salt-bridge cannot form. Another possible explanation for the difference in finger positions are the disulfide bridges present in 9<sup>HN</sup>-7 pol but absent in the Tgo pol structure and in pol  $\beta$  family structures. At least one of the disulfides (Cys428/442) in 9<sup>HN</sup>-7 pol could be directly involved in orienting the fingers relative to the palm (Hopfinger *et al.*, 1999).



ternary pol, DNA and dNTP substrates from the bacteriophage T7 pol ternary complex (Shostak et al., 1998) were modeled into the P<sub>1</sub>N7 pol active site. The model shown in Figure 6 provides further



**Figure 3.** The RNA-binding motif in the NH<sub>2</sub>-terminal domain of 9°N-7 pol. (a) Topology diagram of the complete NH<sub>2</sub>-terminal domain residues 1-228, 338-372. The RNA-binding motif (red) is shown as a helix-loop-helix structure. (b) Sequence alignment of the NH<sub>2</sub>-terminal domain of 9°N-7 pol, RB69 pol, T4 pol, and archaeal pols. Alignment of 9°N-7 pol and T4/RB69 is based upon the crystal structures, and that of 9°N-7 pol and the other archaeal polymerases is based upon sequence alignment of 13 sequences among those considered by Edgell et al. (1997) (data not shown). The archaeal polymerase alignments were performed with the FOLUP algorithm in the GCG package (University of Wisconsin Genetic Computer Group). Secondary structure elements corresponding to 9°N-7 pol are given above the sequence. A consensus sequence was derived for the archaeal polymerases at those positions where at least 70% of the 13 sequences showed the same residue. Residues in yellow are those residues conserved between the archaeal consensus and both bacteriophage T4, RB69 sequences. Position 367 in 9°N-7 pol is entered (see the text for discussion). Abbreviations are as follows: PVL, *Pyrococcus furiosus* 58423; SsKl, *Sulfolobus solfataricus* 18548; MSA, *Methanococcus jannaschii* PRC; Pst, *Pseudomonas aeruginosa* B1. (c) Ribbon representation of the NH<sub>2</sub>-terminal domain of 9°N-7 (left) and RB69 (right) pols. Loop-square C<sup>α</sup> superposition was performed over the region of 9°N-7 pol including strand 4, part of strand 8, and helices E and C, and the domains were separated for clarity. Side chains are shown in green. The long loop between strands 7 and 8 in 9°N-7 pol corresponds to the conformationally variable loop 3 in the bacterial RNAP motif (Sikorski et al., 1997).

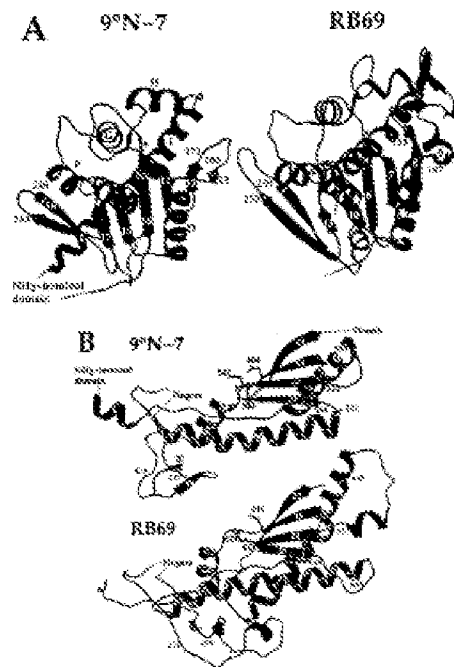
evidence that the position of the fingers in 9°N-7 pol more closely approximates a closed conformation and their position in RB69 pol approximates an open conformation. This model of a ternary complex for a pol  $\alpha$  family polymerase places the dNTP within hydrogen-bonding distance of residues on the fingers D helix that are highly conserved and known by mutagenesis to be functionally important. The corresponding residues on fingers helix E of the RB69 pol are further away and cannot directly interact with dNTP.

The model places residues Y439 and Y484 near the deoxyribose moiety of the incoming dNTP. These residues appear to be functionally analogous to H830 and Y826 of T7 pol, which are responsible for discriminating between deoxy- and ribonucleotides (dNTPs). Y439 is invariant among the pol  $\alpha$  family in the alignment by Brattin et al. (1993) and nearly invariant (one exception) among archaeal pols aligned by Edgell et al. (1997). Mutation of the corresponding residue (Y441) to Val in an exonuclease-deficient *Thermococcus*

*litoralis* (Vent) pol causes a 200-fold loss of discrimination against rNTPs. The aromatic ring appears to be the functionally important moiety, as mutating Y412 in the conserved wild-type discrimination levels (Gardner & Jack, 1999).

Y826 in T7 pol (Y762 in Klenow fragment) has been dubbed the "ribose selectivity site" (Tabor & Richardson, 1990). A Phe residue at this position confers selectivity against incorporation of dinucleoside nucleotides (dNTPs), whereas a Tyr residue in this position allows efficient incorporation of both nucleotide species. The presence of Tyr (Y439) in this position in 9°N-7 pol suggests the ability to incorporate dinucleoside nucleotides, as do Vent (Gardner & Jack, 1999) and human pol  $\alpha$  (Capeless et al., 1992). In fact, Tyr is invariant at this position among the archaeal pols aligned by Edgell et al. (1997), and highly conserved in the pol  $\alpha$  family aligned by Brattin et al. (1993).

The model of a ternary complex with dNTP and DNA places residues M491 and K487 in hydrogen-bonding distance from the triphosphate moiety of



**Figure 4.** Comparisons of 9°N-7 and RB69 pols in different conformations to indicate loop segments that are shorter in 9°N-7 pol. Left: superposition of 9°N-7 pol. Loop segments of 9°N-7 pol were performed over the region in blue, and the domains were separated for side-by-side comparison. Loop regions are shown in magenta and their residue endpoints are numbered. (A) Comparison of the exonuclease domains. Indicated with purple asterisks are the active site carboxylates mutated to Ala in the case of the 9°N-7 pol used in this study. (B) Comparison of the palm domains. The three active-site carboxylate groups are depicted with side-chains.

the incoming dNTP. Both of these residues are invariant in the pol  $\alpha$  family (Brattwalle & Ito, 1993), and nearly invariant (one exception) among archaeal pols (Edgell *et al.*, 1997). Mutation of the corresponding residues (G498, K498) in Vent (exo-) pol severely decreases enzyme activity (Cardiner & Jack, 1999).

#### Concerted domain movement

The difference in position of the fingers subdomain in 9°N-7 and RB69 pols is part of a larger conformational change involving the 3'-5' exonuclease and NH<sub>2</sub>-terminal domains. Comparing these two pol structures shows that in one of the pairs, an essentially rigid-body rotation has occurred involving three of the five subdomains. This concerted movement affects both the position of the fingers relative to the pol active site (open

tense closed conformation), as well as the position of the exonuclease active site relative to the pol active site. The 9°N-7 and RB69 pol structures may approximate different states along the reaction pathway corresponding to DNA synthesis and 3'-5' exonuclease proofreading activities.

When these two polymerases are aligned in the palm (the blue region in Figure 4B), the exonuclease and fingers are displaced between the proteins (Figure 5A). If the enzymes are aligned in the exonuclease domain (see Figure 5A), the fingers superimpose almost exactly (Figure 5B). Moving from a palm to an exonuclease-based alignment also brings the first module (residues 1-31) of the NH<sub>2</sub>-terminal domain into identical positions (not shown). The joint motion of the first NH<sub>2</sub>-terminal module and the exonuclease may reflect the need to maintain ionic networks at the interface. There are two five-membered ionic net-

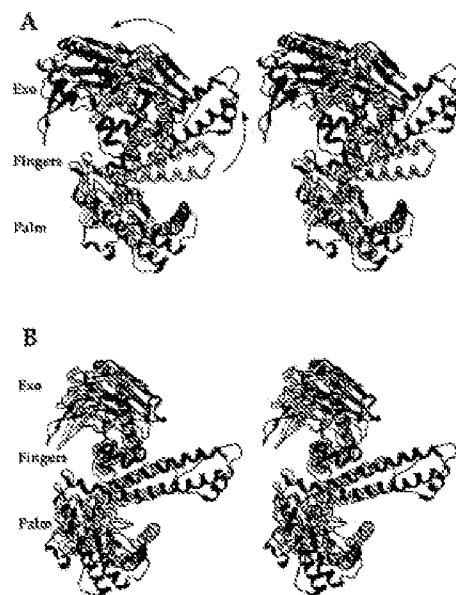


Figure 5. Least-squares  $C\alpha$  superpositions of 9°N-7 and R369 pols in the (a) palm subdomain or (b) exonuclease domain. The 9°N-7 pol backbone is shown in yellow, and its active-site nucleobase groups in gold. The R369 pol backbone is shown in green, and its active-site residues in magenta. The central 3'-phases of the exonuclease domain is light blue (9°N-7 pol) or dark blue (R369 pol) to allow tracking of the domain motion. The precise regions used in the palm and exonuclease superpositions are shown in Figure 4. The N15-terminal domain has been omitted for clarity. Arrows in (a) indicate the direction of fingers and exonuclease movement when moving from (a) to (b).

works formed between the first module and exonuclease (Figure 7). In addition, a three-membered network is formed between the third NH<sub>2</sub>-module (R346) and the exonuclease (Figure 7). This network is conserved among nearly all archaeal pols (Edgell *et al.*, 1997), but none is present in R369 pol.

Comparison of the Tgo pol structure (Kaphner *et al.*, 1999) with that of 9°N-7 and R369 pols using palm and exonuclease-based superpositions gives results similar to those in Figure 5, providing further support for the notion of a concerted domain movement.

A model was constructed for the R369 pol (Wang *et al.*, 1997) showing how substrate dNTPs could shuttle between the pol and exonuclease active sites. When 9°N-7 and R369 pols are aligned in the palm, the exonuclease active site in the former is tilted out and away from the pol active site, making it impossible for the DNA to shuttle. The exonuclease position in R369, but not that in 9°N-7 pol, is therefore consistent with an editing conformation. It is interesting that this confor-

mation also means that the fingers are not in position to bind dNTP (see above). Taken together, these considerations suggest that during the replication cycle of family II pols, there is concerted movement of the exonuclease, NH<sub>2</sub>-terminal domain, and fingers relative to the catalytic region of the palm.

This concerted movement may be the structural basis for the functional coupling of polymerase and exonuclease domains, which is unique in the pol  $\alpha$  family. In this family it is possible to generate site-directed mutations in one domain that exert an indirect, negative effect on the other (Reto-Kraus & Klenow, 1993; Abdus-Sattar *et al.*, 1996). This contrasts with pol I pols like E3, where these activities are completely confined to their respective domains (Olin *et al.*, 1985).

#### Molecular basis of thermostability

*Thermococcus* sp. 9°N-7 grows at temperatures of 84–90°C, and its pol has a temperature optimum of 70–80°C (Paele *et al.*, 1996). It has a half-life of 6.7



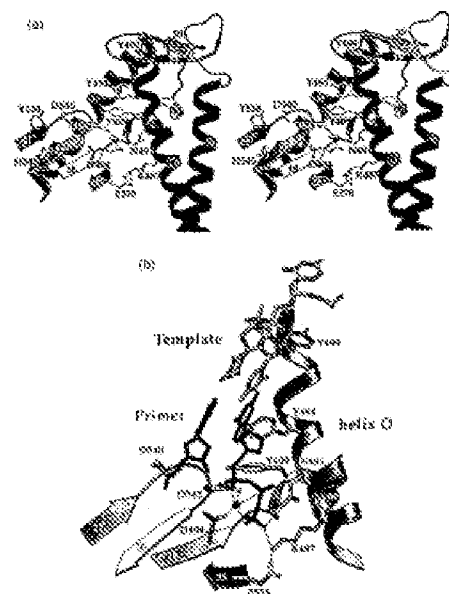


Figure 6. The active site of 9°N-7 pol and a model of a ternary complex. (a) Stereoview of the active site. Residues with indicated side-chains are discussed in the text. Hydrogen bonds are as broken lines, and the two disulfide bridges are shown as solid lines. K487 in this structure is involved in a salt-bridge with E579 of the palm. (b) Model of a ternary complex of 9°N-7 pol. For clarity only the incoming base and the first primer-strategy base-pairs are shown. Hydrogen bonds are shown as broken lines, and non-covalent interactions are indicated as green spheres. The 9°N-7 pol and 27 pol ternary complex (Doublié *et al.*, 1998) were superimposed to the palm (65% Å root for C $\alpha$  atoms). The minor conformation was adjusted for D542 and D468 in 9°N-7 pol, and the 3' arm including D542 was tilted downward, in a manner analogous to that observed between the open-type and binary complex structure of *Saccharomyces cerevisiae* pol (Kiefer *et al.*, 1997, 1998).

hous at 95°C (R.6. *Saccharomyces cerevisiae*), whereas *Thermus aquaticus* (Taq) DNA pol has a half-life of 1.6 hours at 95°C (Hong *et al.*, 1993). The structure of 9°N-7 pol indicates a few key strategies for this hyperthermostability, some of which appear general to archaeal DNA pols.

A surprising feature of the 9°N-7 pol is that it contains two disulfide bridges (Figures 1(a) and 6(a)). The potential for the same bridges in form was also observed in Tgo pol (Hopwood *et al.*, 1999). Although not normally the case in Bacteria or Eucarya, an increasing number of cytosolic proteins with disulfide bridges are being discovered in the Archaea (DeClerck *et al.*, 1996; Singley *et al.*, 1998). The stabilizing role of disulfide bridges has been well documented (Calkins *et al.*, 1994; Cooper *et al.*, 1993). Introduction of disulfide bridges therefore appears to be a common strategy for archaeal protein stability.

Alignment of a large number of archaeal pols (Edgell *et al.*, 1997) suggests that having at least one of these disulfides is important for their thermostability. In fact, the two-stranded  $\beta$ -sheet

encompassing C442 corresponds to sequence motif D in archaeal pols (Edgell *et al.*, 1997). Based on whether Cys is present in the corresponding positions, all the pols discussed by Edgell *et al.* (1997) are predicted to have at least one of the two disulfide bridges seen in 9°N-7 pol, with the exception of *M. maris* and *S. shibatae* B3 pols. The thermostability of *M. maris* pol may be partly caused by a lack of disulfide bridges. The *S. shibatae* B3 pol, like the *S. solfataricus* P2 B3 pol, is highly divergent in sequence from other archaeal pols, and it is unclear whether either of these functions *in vivo* (Edgell *et al.*, 1997).

An increased number of salt-bridges relative to mesostable homologs is often cited as a determinant of protein thermostability (DeClerck *et al.*, 1996; Kornacker *et al.*, 1995; Chan *et al.*, 1998; Henning *et al.*, 1995). The 9°N-7 pol shows a substantial increase in the fraction of charged residues participating in salt-bridges (47%) compared with R689 pol (39%). These results are similar to a thermostability study of *Pancreaticus ferrous* glutamate dehydrogenase (Yip *et al.*, 1995). The addition of

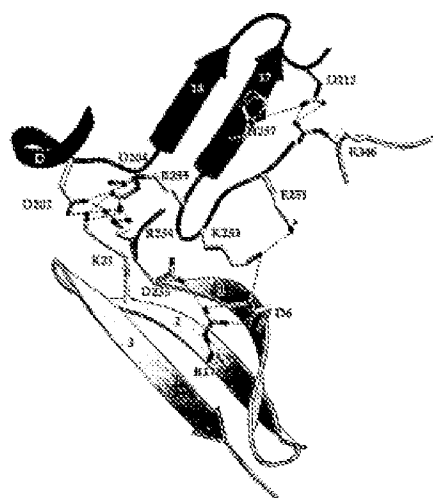


Figure 7. The extensive ionic networks at the interface of the NH<sub>2</sub>-terminal and 3'-5' exonuclease domains.

that study found a marked preference for Arg residues in the ionic interactions of the thermostable enzyme, but no such preference is evident here. The same fraction (38%) of Arg residues is used in ionic interactions in both 9°N-7 and RB69 pols, whereas a much higher proportion of Glu residues participate in salt-bridges in the 9°N-7 pol (83%) compared with RB69 pol (33%).

The number and distribution of salt-bridges within domains does not substantially differ between 9°N-7 and RB69 pols. At the interfaces between subdomains, however, the differences in ionic networks are striking. The proportion of ionic interactions at interfaces in the 9°N-7 pol (21%) is over twice that in RB69 pol (9%). The differences lie at the interface of the exonuclease domain with the NH<sub>2</sub>-terminal domain (Figure 7), and at the interface of the exonuclease with the thumb, where a two and a three-member ionic network occur in 9°N-7 pol (compared with none in RB69 pol (not shown)).

Burial of the charged termini of proteins has been cited as another factor that can confer thermostability (Moras *et al.*, 1995). The NH<sub>2</sub>-terminal methionine (M1) of 9°N-7 pol is stabilized by a hydrophobic cluster formed by L125, F327, R255, V206, and L341 while the corresponding residue of RB69 pol is completely exposed to solvent. The *B*-factor for the C $\alpha$  of M1 in 9°N-7 pol is 26 Å<sup>2</sup>,

whereas for M1 in RB69 pol, it is 95 Å<sup>2</sup>. While burial of the N terminus may be important for the thermostability of the 9°N-7 pol, the same does not hold for the C terminus. The last 25 residues are not visible in the electron density, similar to the case of RB69 pol. The solvent accessibility of the C terminus of these pols may reflect the need for this region to interact with a processivity accessory protein, which is known to be the case in the T3 replication complex (Berdic *et al.*, 1996).

Another common strategy for protein thermostability is to lower the solvent-accessible surface area of the protein and to increase the proportion of buried structure (Kamat *et al.*, 1995; Chan *et al.*, 1995). This translates into a more compact structural design. There are at least 12 examples of loop segments in RB69 pol that are much shorter or absent in 9°N-7 pol. Some of the more striking examples are shown in Figure 8. Alignment of 16 archaeal pols (Edgell *et al.*, 1997) indicates that they share practically all of these sequence "deletions". The T3 pol structure also revealed shortened loop segments relative to RB69 pol (Hopfner *et al.*, 1999). Nevertheless, the molar ratio of solvent-accessible surface area to volume for both 9°N-7 and RB69 pols is the same (0.33). Thus, while lowering the surface area to volume ratio is a common strategy for thermostability, it is not the primary basis for the stability of 9°N-7 pol.

## Materials and Methods

### Purification, crystallization, and data collection

*Thermococcus* sp. 9°N.7 polymerase (wild-type and the D141A,D143A, exonuclease-deficient mutants) was over-expressed and purified as described (Scott-Smith et al., 1995). Crystallization, cryoprotection, data collection and reduction of native crystals are described (Zheng et al., 1998). Derivatives were prepared by soaking native crystals in solubilization solutions (Zheng et al., 1998) supplemented with 22.7 mM sodium dithionite, 10 mM dithiothreitol (Biomol) for 31 days (thimerosal-1), 3.0 mM  $K_2P_2O_7$  for one hour (PDC-1), 1.5 mM d-p-tetradecyl methylphosphonotriethylammonium (d) cistate (PIP) for one day (PIP-1), or 1.0 mM Baker's monocalcium for 30 hours (BAH). These crystals were stopped through solubilization solution containing 8% (v/v) sucrose, 16% (v/v) monomers, and 20% sucrose (one to five hours). Additional derivatives were collected with the improved cryoprotocols procedure reported (Zheng et al., 1998) by soaking native crystals in 23.0 mM thimerosal for 8.3 days (thimerosal-2), 3.0 mM  $K_2P_2O_7$  for seven days (PDC-2), and 1.5 mM PIP for 38 hours (PIP-2).

### Structure Determination

The structure of the D141A,D143A mutant of 9°N.7 polymerase was determined by the method of multiple isomorphous replacement (MIR). A number of native and derivative crystals were used to solve the structure because of problems with non-isomorphism (Table 1). Three native datasets were collected from single crystals. NAT-1 was mounted in the liquid nitrogen stream directly from cryoprotectant, whereas NAT-2 and -3 were flash-frozen in liquid nitrogen prior to mounting. The crystals belong to space group  $P2_12_12_1$  with unit cell dimensions of approximately  $a=96.3$  Å,  $b=103.3$  Å,  $c=112.2$  Å (for NAT-3). One molecule is present per asymmetric unit, giving a solvent content of approximately 60%.

A difference Patterson map of thimerosal-1 was calculated using the program FFT in the CCP4 suite (CCP4, 1994). The heavy-atom site for this derivative was identified with the program RSPG (Knight, 1988). This site was used to calculate initial phases for NAT-1 at 5 Å resolution using the program MLPHARE (Otterwinowski, 1991). Difference Fourier synthesis with the initial phases revealed three sites for the PDC-1 derivative. Two more sites for this derivative were discerned with the phases derived from both thimerosal-1 and PDC-1. The correct handedness of the phasing information from thimerosal-1 was determined using MLPHARE, and subsequent scattering data from the derivatives were included in the phase calculation. Three sites for the BAHg derivative and four sites for PIP-1 were obtained from difference Fourier maps calculated to 5 Å resolution. All of these heavy-atom sites were included in subsequent phase calculations with NAT-1. The high-resolution phasing limit was extended to 3.5 Å. Because of the high solvent content in the crystals, use of the solvent-flattening program DM (Cowtan, 1994), in combination with histogram matching, improved the phase substantially. A polyatomic model was built into the improved electron density map of NAT-1 with the program O Jones & Kjeldgaard, 1993) and refined by the program X-PLOR (Brünger, 1992). Phase combination using the program SHELX (Sheldrick, 1986) further improved the map during building and refinement.

Identification of side-chain densities was possible only after collecting a higher-resolution native dataset (NAT-2) along with diffraction data for three more derivatives obtained under improved cryoprotection conditions (thimerosal-2, new site; PDC-2, four sites; PIP-2, five sites). These derivatives were used to calculate MIR phases of NAT-2 at 3.0 Å resolution. Partial model phases of NAT-1 were calculated using the refined polyatomic model derived from NAT-1. Because of significant differences in unit cell dimensions between NAT-1 and 2, it was first necessary to subject NAT-2 to rigid-body refinement against NAT-1 in X-PLOR. Combination of the polyatomic model phases and MIR phases with SHELX, improved the electron density map. Model building, refinement, and phase combination were continued until a complete polyanionic model could be built in the final stage of refinement. NAT-3 was used to extend the resolution limit to 2.1 Å and water molecules were added.

### Coordinate files and illustrations

The *Thermococcus* sp. 9°N.7 polymerase atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank under the accession code 1U8H. The RBM coordinates used for comparison in the microscopy are those of the orthorhombic crystal form (accession code 1WAB). Figures were prepared within the RSC Shownote program (Shown Graphics, Inc.) entirely (1(a), 2, 3(a) and 3(b)) or with images imported from MOLSCRIPT (Merritt, 1991) or SETOR (Jin, 1997) (Jones, 1995).

## Acknowledgments

We thank New England Biolabs, Inc. for their collaboration on this project especially S. Kay Williams and Rebecca Korman for protein purification and Francine Parker and William Jack for helpful discussions. We thank Anne McCoy, David Owen, Daniela Stock, Jean-Luc Jost, and Monique Heflinger for critical comments on the manuscript. We also thank Anne McCoy, Jeff Taylor, and Scott Johnson for assistance in figure preparation. This work was supported by grants to L.S.B. from the ACS (315-823-A), the North Carolina Biotechnology Center, and the Leslie Schukert Program.

## References

- Abdus, Satter A. K., Lin, T. C., Jones, C. & Kungberg, W. H. (1992). Functional heterogeneity and exonuclease kinetic parameters of point mutations in bacteriophage T4 DNA polymerase. *Biochemistry*, **31**, 16021-16025.
- Barnes, S. M., Dubrovich, C. E., Palmer, J. D. & Pace, N. S. (1994). Perspectives on archaeal diversity, taxonomy, and ecologically based environmental DNA sequence. *Proc. Natl. Acad. Sci. USA*, **91**, 8188-8195.
- Brown, L. S. & Smith, T. A. (1991). Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.*, **10**, 25-33.
- Bentley, A. J., Soumloff, P. & Beckett, S. J. (1990). The carboxyl terminus of the bacteriophage T4 DNA

- polymerase is required for heterozygote complex formation. *Proc Natl Acad Sci USA*, 93, 12822-12827.
- Birnst, L., Bernad, A. & Salas, M. (1992). Evidence favoring the hypothesis of a conserved 5'-3' exonuclease active site in DNA-dependent DNA polymerases. *Gene*, 112, 139-144.
- Brenishe, D. S. & Ro, J. (1993). Comparison, alignment, and phylogenetic relationships of DNA polymerases. *Mol. Acids Res.* 21, 787-802.
- Bringer, A. T. (1992). X-PLOR Version 3.1: A System for X-ray Crystallography, and 1992, Yale University Press, New Haven, CT.
- Rütt, C. J., White, D., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. C., Blake, J. A., Findorff, L. M., Clayton, R. A., Gocayne, J. D., Kaulavage, A. S., Shueherty, B. A., Tomb, J. P., Adams, M. D. et al. (1990). Complete genome sequence of the mesophilic archaeon, *Methanocaldococcus jannaschii*. *Science*, 257, 185-187.
- Rosengart, C. A. & Steitz, T. A. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrates. *Curr Opin Struct Biol*, 8, 34-43.
- Rosini, C. C. & Freyberger, C. (1994). Conserved structure and diversity of functions of DNA-binding proteins. *Science*, 265, 815-821.
- Chou, M. K., Makous, S., Klotz, A., Adams, M. W. & Roen, D. C. (1993). Structure of a hyperthermophilic triphosphatase, alkaline ferredoxin, oxalomalate. *Science*, 262, 1465-1469.
- Collaborative Computational Project No. 4 (1994). The CCP4 suite programs for protein crystallography. *Acta Crystallogr. sect. D*, 50, 216-233.
- Cropper, A., Eyles, S. L., Rodford, S. E. & Dobson, C. M. (1982). Thermodynamic consequences of the removal of a disulfide from hen lysozyme. *J. Mol. Biol.*, 158, 429-443.
- Copeland, W. C., Lam, N. K. & Wang, T. S. F. (1993). Fidelity studies of the human DNA polymerase  $\alpha$ : the most conserved region among  $\alpha$ -like DNA polymerases is responsible for metal-induced inhibition in DNA synthesis. *J. Biol. Chem.* 268, 11981-11989.
- Copeland, W. C. & Wang, T. S. F. (1993). Mutational analysis of the human DNA polymerase  $\alpha$ : the most conserved region in  $\alpha$ -like DNA polymerases is involved in metal-specific catalysis. *J. Biol. Chem.* 268, 11978-11980.
- Crowson, N. (1994). "DNA" an antineoplastic prodrug for gene improvement by density modification. *Acta Cryst. and ESP-EACRIS News. Protein Crystallogr.* 21, 34-38.
- DeGeeck, B. S., O'Brien, S., Fleming, P. J., Gager, J. H., Jackson, S. P. & Rigby, P. B. (1996). The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J. Mol. Biol.* 264, 1022-1034.
- Delarue, M., Poch, O., Tordo, V., Moras, D. & Argos, P. (1990). An attempt to unify the structure of polymerases. *Protein Eng.* 4, 461-467.
- DeLong, E. S., Wu, E. Y., Poczobski, B. S. & Jovin, R. V. M. (1994). High abundance of Anabaena in Antarctic marine picoplankton. *Science*, 264, 695-697.
- Herbst, V., Pinnerstein, J. K. & Jovin, C. M. (1993). Structure-function analysis of 5'-3' exonuclease of DNA polymerases. *Molecular Biotech.* 5, 383-385.
- Doublie, S., Tabor, S., Lang, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication enzyme at 2.2 Å resolution. *Nature*, 391, 251-258.
- Doublie, S., Sawaya, M. R. & Ellenberger, T. (1999). An open and closed case for all polymerases. *Structure*, 7, 831-835.
- Edgell, D. R. & Madeline, W. F. (1997). Archives and the origins of DNA replication proteins. *Cell*, 89, 989-998.
- Edgell, D. R., Klock, H. P. & Madeline, W. F. (1997). Gene duplication in evolution of archaeal family B DNA polymerases. *J. Bacteriol.* 179, 2657-2660.
- Eum, S. H., Wang, J. & Roeder, T. A. (1996). Structure of  $\beta$  polymerase with DNA at the polymerase active site. *Nature*, 382, 278-281.
- Evans, S. V. (1993). GEMO: hardware-lighted three-dimensional unit model representations of macromolecules. *J. Mol. Graphics*, 11, 134-138.
- Garabito, V. & Pharo, W. (1998). Cytotoxicity, mutagenesis, and mechanisms of action of 2,6-dichloro-4-nitroguanine in Chinese hamster ovary cells. *Cancer Res.* 58, 1517-1524.
- Gardner, A. R. & Jack, W. E. (1998). Determination of nucleotide sugar recognition in an archaeal DNA polymerase. *Biochem. Acids Res.* 25, 2345-2353.
- Geddes, R. S., Agarwal, S., Francis, V. A., Sami, D. V. & Salas, P. (1998). Thermal stabilization of hybridization synthesis by engineering non disulfide bridges across the dimer interface. *J. Mol. Biol.* 285, 89-94.
- Harzer, M., Meyer, C., Kötter, C., Guldert, G., Götner, P. & Pöschel, W. (1994). Antigenic recognition regulation of the ribosomal M61L gene in the archaeobacterium *Methanocaldococcus jannaschii*. *J. Bacteriol.* 176, 839-848.
- Hwang, M., Darmon, S., Berner, S., Kirschner, K. & Janszewska, J. H. (1995). 2.8 Å structure of nucleic acid triphosphatase synthase from the hyperthermophilic *Sulfolobus solfataricus*: possible determinants of protein stability. *Structure*, 3, 1285-1296.
- Hopfinger, S. P., Schlegel, A., Singh, R. A., Lase, F., Anandaram, W., Huber, R. & Angerer, B. (1995). Crystal structure of a thermophilic type B DNA polymerase from *Thermococcus goniorum*. *Proc. Natl. Acad. Sci. USA*, 92, 3605-3609.
- Huang, H., Cheng, S., Verdine, G. L. & Harrison, S. C. (1998). Structure of a reversibly trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, 282, 1665-1675.
- Huang, P. & Pharo, W. (1995). Fluorescence- and gamma-ray-induced apoptosis incorporation of analogs into DNA is a critical event. *Cancer Chemother. Pharmacol.* 36, 181-188.
- Jones, T. A. & Kjeldgaard, M. (1999). O Version 3.0: The Atomic, Crystallographic, Lippstadt, Sweden.
- Joyce, C. M. & Sharp, T. A. (1994). Function and Structure relationships in DNA polymerases. *Annu. Rev. Biochem.* 63, 777-822.
- Kabat, W. & Borek, C. (1963). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biochemistry*, 22, 2577-2637.
- Keating, M. J., McCredie, K. B., Boey, G. P., Smith, T. L., Gehring, S. & Freireich, R. J. (1982). Improved prognosis for long-term survival in adults with acute myelogenous leukemia. *J. Am. Med. Assoc.* 248, 2881-2886.
- Kiefer, J. S., Moss, C., Horne, C. J., Richardson, S. L., Hargreave, H. H., Beeman, J. C. & Bessie, L. S. (1997). Crystal structure of a thermophilic *Bacillus* DNA polymerase I large fragment at 2.1 Å resolution. *Structure*, 5, 95-108.

- Kiefer, J. R., Maci, C., Beeson, J. C. & Beese, L. S. (1998). Visualizing DNA replication in a catalytically active *Nautilus* DNA polymerase crystal. *Nature*, **391**, 524-527.
- Kim, V., Eom, S. H., Wang, J., Lee, G. S., Suh, S. W. & Seitz, T. A. (1995). Crystal structure of *Thermococcus aquaticus* DNA polymerase. *Nature*, **376**, 642-646.
- Kong, H., Rocco, R. B. & Jack, W. T. (1993). Characterization of a DNA polymerase from the hyperthermophilic *Archaea Thermococcus litoralis*. *J. Biol. Chem.* **268**, 1965-1975.
- Korenchuk, I., Sjöberg, H., Huber, R., Tomach, A. & Janszko, B. (1995). The crystal structure of tetrahydroaldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. *J. Mol. Biol.* **246**, 521-531.
- Kumita, S., Nagai, M., Barnes, W. M., De Luca, E. & Vakserman, G. (1995). Crystal structure of the large subunit of *Thermus aquaticus* DNA polymerase I at 2.5 Å resolution. *Proc. Natl Acad. Sci. USA*, **92**, 9254-9258.
- Li, Y., Krumm, S. & Wosman, G. (1996). Crystal structure of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I structured basic for nucleotide incorporation. *EMBO J.* **15**, 7514-7527.
- Olso, D. L., Brick, P., Kariolis, R., Kung, W. G. & Steitz, T. A. (1995). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTTP. *Nature*, **375**, 752-755.
- Owczonicki, Z. (1991). In *Nonprotein Replicases and Anomalous Scattering* (Wol, W., Kratoch, P. R. & Leslie, A. C. W., eds), vol. 80. Science and Engineering Research Council, Washington, UK.
- Pfisterer, W., Sawaya, M. R., Kistele, A., Wilson, S. H. & Kowd, J. (1994). Structure of binary complexes of rat DNA polymerase beta, a DNA template-primer, and dCTP. *Science*, **264**, 1930-1935.
- Perle, F. B., Komer, S. & Kong, H. (1996). Thermostable DNA polymerases. *Advan. Protein Chem.* **48**, 397-435.
- Priestle, J. E. (1991). A program to produce both detailed and schematic drawings for protein structures. *J. Appl. Crystallog.* **24**, 846-850.
- Rued, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallog.* **42**, 140-149.
- Ross, D. C. & Adams, M. W. W. (1993). Hyperthermophiles: taking the heat and living it. *Structure*, **1**, 283-295.
- Rob-Kozlowski, L. J. & Moras, R. J. (1995). Genetic and biochemical studies of bacteriophage T4 DNA polymerase 2'-5' exonuclease activity. *J. Biol. Chem.* **270**, 12710-12718.
- Robertson, L. E. & Florkett, W. (1992). High-dose cytosine arabinoside in chronic lymphocytic leukemia: a clinical and pharmacologic analysis. *Leuk. Lymphoma*, **10**, 45-48.
- Sharon, Y., Kravitz, U., Riss, L. M., Williams, K. R. & Seese, L. S. (1997). Crystal structure of the two DNA binding domains of human hgp32p A1 at 1.77 Å resolution. *Nature Struct. Biol.* **4**, 215-222.
- Shimrin, L. C. & Dennis, P. F. (1989). Characterization of the L1, L2, L3, and L4 regulatory ribosomal protein gene cluster of the halophilic archaeobacterium *Haloferax volcanius*. *EMBO J.* **8**, 1225-1235.
- Singh, M. R., Kuper, M. N. & Lohmeyer, J. A. (1999). X-ray structure of pyrimidine catalytic peptide from the hyperthermophilic archaeon *Thermococcus litoralis*. *Structure*, **7**, 237-244.
- Smithworth, M. W., Kong, H., Rocco, R. B., Ware, J., Janszko, B. W. & Parker, F. B. (1996). Cloning of thermostable DNA polymerases from hyperthermophilic marine *Archaea* with emphasis on *Thermococcus* sp. F-102 and mutants showing 5'-3' exonuclease activity. *Proc. Natl Acad. Sci. USA*, **93**, 6265-6268.
- Takao, S. & Richardson, C. C. (1995). A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between decay- and dideoxycytosine nucleotides. *Proc. Natl Acad. Sci. USA*, **92**, 6339-6343.
- Towle, C., Biddy, B., Parsons, D. & Cold, L. (1996). Autogenous translational operator recognized by bacteriophage T4 DNA polymerase. *J. Mol. Biol.* **259**, 744-751.
- Wang, L., Yu, P., Liu, Y. C., Kung, W. G. & Steitz, T. A. (1996). Crystal structure of an N15-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with double-stranded DNA. *Biochemistry*, **35**, 8118-8128.
- Wang, J., Sakai, A. K. M. A., Wang, C. C., Korman, J. D., Kung, W. G. & Steitz, T. A. (1997). Crystal structure of a pol  $\alpha$  family replication DNA polymerase from bacteriophage RB69. *Cell*, **89**, 1067-1080.
- Wiese, C. R., Kandler, O. & Winkler, M. L. (1996). Towards a natural system of organisms: Proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc. Natl Acad. Sci. USA*, **93**, 4216-4221.
- Wong, S. W., Wong, A. F., Yoon, P. M., Anz, N., Peterson, B. E., Aris, N., Korn, D., Harkapiller, M. W. & Wang, T. S. E. (1996). Human DNA polymerase  $\alpha$  gene expression is cell proliferation dependent and its primary structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. *EMBO J.* **15**, 37-47.
- Yip, S. F., Shillman, T. J., Bhatia, K. L., Argyropoulos, P. J., Bhatia, P. J., Schickel, S. R., Engel, P. C., Pappas, A., Choudhury, R., Consolvi, V., Srinivasan, S. & Ren, D. W. (1995). The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a very role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure*, **3**, 1347-1358.
- Zhou, M., Mao, C., Rodriguez, A. C., Kider, J. R., Rocco, R. B. & Seese, L. S. (1996). Crystallization and preliminary diffraction analysis of a hyperthermostable DNA polymerase from a *Thermococcus* archaeon. *Acta Crystallog.* **42**, 944-949.

Edited by D. Rees

(Received 15 June 1999; received in revised form 24 March 2000; accepted 24 March 2000)

## Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase\*

Received for publication, July 29, 2003, and in revised form, December 23, 2003  
Published, JBC Papers in Press, December 29, 2003, DOI 10.1074/jbc.M309382000

Andrew F. Gordon<sup>†</sup>, Catherine M. Joyce<sup>‡</sup>, and William E. Jack<sup>§</sup>

From <sup>†</sup>Blue Bird Biological Inc., Beverly, Massachusetts 01915 and the <sup>‡</sup>Department of Molecular Biophysics and <sup>§</sup>Biochemistry, Yale University, New Haven, Connecticut 06510

Comparative kinetic and structural analyses of a variety of polymerases have revealed both common and divergent elements of nucleotide discrimination. Although the parameters for dNTP incorporation by the hyperthermophilic archaeal Family B Vent DNA polymerase are similar to those previously derived for Family A and B DNA polymerases, parameters for analog incorporation reveal alternative strategies for discrimination by this enzyme. Discrimination against ribonucleotides was characterized by a decrease in the affinity of NTP binding and a lower rate of phosphoryl transfer, whereas discrimination against ddNTPs was almost exclusively due to a slower rate of phosphodiester bond formation. Unlike Family A DNA polymerases, incorporation of 8-(p-hydroxyphenylthio)-dGTP (HPT-dGTP) (where X is adenine, cytosine, guanine, or thymine, respectively) by Vent DNA polymerase was enhanced over ddNTPs via a 50-fold increase in phosphoryl transfer rate. Furthermore, a mutant with increased propensity for nucleotide analog incorporation (Vent<sup>108S</sup> DNA polymerase) had unaltered dNTP incorporation while displaying enhanced nucleotide analog binding affinity and rates of phosphoryl transfer. Based on kinetic data and available structural information from other DNA polymerases, we propose active site models for dNTP, ddNTP, and rNTP selection by hyperthermophilic archaeal DNA polymerases to rationalize structural and functional differences between polymerases.

All free-living organisms encode several DNA polymerases that are primarily responsible for the replication and maintenance of their genomes, thereby ensuring accurate transmission of genetic information (1–3). The variety of identified DNA polymerases can be classified into Families A, B, C, and X according to amino acid sequence similarities to *Escherichia coli* polymerases I, II, III, and IV, respectively (4, 5). Additional families have been identified, including the two-subunit replicative DNA polymerases from hyperthermophilic Archaea (Family D) (6) and eukaryotic DNA polymerase  $\beta$  and terminal transferase (Family X) (4).

Structural and kinetic analyses of Family A (7–14) and Family B (15–25) DNA polymerases have increased the understanding of nucleotide selection and incorporation mechanisms. Al-

though amino acid sequences diverge between these two families, the structures of Family A and B DNA polymerases show recognizable finger, thumb, and gate subdomains that allow recognition of structural elements important for function (8, 11). In the case of Family B DNA polymerases from bacteriophage T7, *Escherichia coli* (Klenow fragment), large fragment of DNA polymerase  $\beta$ , and *Thermosta aquaticus*, as well as the Family B DNA polymerase from bacteriophage B299, incorporation of the structural information is complemented by steady-state and pre-steady-state kinetic studies, allowing a detailed description of the polymerization pathway. Reaction parameters describing the discrimination against naturally occurring nucleotide analogs incorporated *in vivo*, such as dNTPs or structural nucleotide analogs, such as ddNTPs and d-labeled ddNTPs (18, 26–30), have added insights into the basis for nucleotide discrimination.

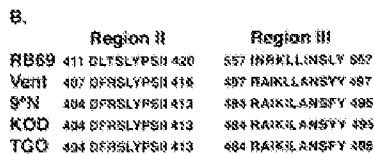
Hyperthermophilic archaeal DNA polymerases have not been scrutinized in such detail, hampering a complete characterization and comparison with other polymerases. Family B DNA polymerases from hyperthermophilic Archaea *Thermococcus* sp. 97N (22), *Thermococcus gorgonarius* (18), and *Pyrococcus kodakarensis* KOD1 (24) and mesophilic bacteriophage RB69 (23) have high sequence and structural homologies and provide a framework for analysis of active site structure and function in this enzyme family (Fig. 1). Furthermore, steady-state kinetic studies have identified hyperthermophilic DNA polymerase residues important for polymerization and exonuclease activities and for nucleotide binding (18, 20, 31–35). Nucleotide analogs have also been important in identifying dNTP recognition determinants important in the polymerase reaction (32–39) and have proven useful in a variety of other biology applications, such as DNA sequencing and detection of single nucleotide polymorphisms (37–41). The group of analogs, 5'-[3'-hydroxyphenylthio]dNTPs (HPT-dNTPs) (where X is adenine, cytosine, guanine, or thymine, respectively) is particularly intriguing due to the wide spectrum of incorporation efficiencies noted in different DNA polymerases, even within the same family of polymerase. For example, within Family B, the herpes simplex virus type 1 and human cytomegalovirus DNA polymerases incorporate rNTPs more efficiently than dNTPs, whereas human polymerase  $\beta$  more readily inserts dNTPs over rNTPs (42). Such differences have been exploited in drug therapies where infective agents encode polymerases that more readily insert rNTP than does the host DNA polymerase (43). Hyperthermophilic archaeal DNA poly-

\* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Supported by National Institutes of Health Grant GM-60000.

§ To whom correspondence should be addressed. Present address: Blue Bird Inc., 22 Tower Rd., Beverly, MA 01915. Tel.: 978-681-0964; Fax: 978-681-1380; E-mail: jack@bluebird.com.

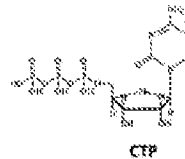
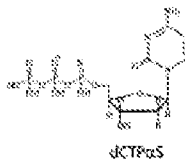
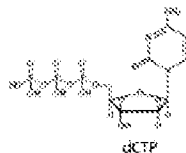
The abbreviations used are: dNTP, 2'-deoxyribosyl-5'-triphosphate; rNTP, 2'-ribosyl-5'-triphosphate; HPT-dNTP, 5'-[3'-hydroxyphenylthio]dNTP; HPT-dGTP, 5'-[3'-hydroxyphenylthio]dGTP; HPT-dATP, 5'-[3'-hydroxyphenylthio]dATP; HPT-dCTP, 5'-[3'-hydroxyphenylthio]dCTP; HPT-dTTP, 5'-[3'-hydroxyphenylthio]dTTP; HPT-dGTP, 5'-[3'-hydroxyphenylthio]dGTP; HPT-dATP, 5'-[3'-hydroxyphenylthio]dATP; HPT-dCTP, 5'-[3'-hydroxyphenylthio]dCTP; HPT-dTTP, 5'-[3'-hydroxyphenylthio]dTTP.



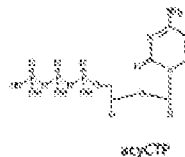
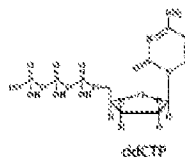
Although instructive, these steady-state observations failed to address the underlying kinetic mechanism responsible for nucleotide and nucleotide leaving group incorporation in hyperthermophilic DNA polymerases. Therefore, we initiated pre-steady-state kinetic studies to measure the kinetic mechanism of nucleotide discrimination in *Ves* and other DNA polymerases.

Conversely, the fluorescently labeled DNA primer template is present was monitored by denaturing PAGE and autoradiated thymidine deoxythymine nucleotides. Product DNA was denatured by mixing a 10  $\mu$ l aliquot of quenched sample with 40  $\mu$ l of formamide and 5.0 mM EDTA, set boiling at 100°C for 2 min. Fluorescent 5'-PAM-labeled 25-nter oligonucleotide template and 8-7-AAM-labeled 30-nter oligonucleotide product bands were fractionated by electrophoresis on 8% 5' deoxyribose and 10% polyacrylamide denaturing gel using an AFS-3000 automated nucleic acid sequencer (Applied Biosystems, Foster City, CA). The sequencing reaction was performed in 10  $\mu$ l of sequencing buffer (Applied Biosystems). The final order was confirmed by independent additional analysis of each sequencing concentration was estimated from a plot of fluorescence intensity. Data sets acquired at 100  $\mu$ M were subsequently plotted on a Scatchard plot.

## A. Nucleotides



## B. Nucleotide terminators



## C. Dye-nucleotide terminators

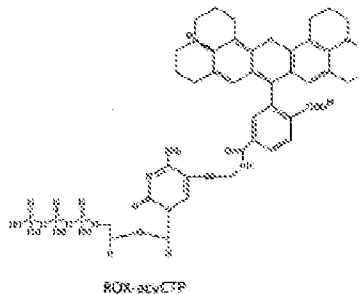
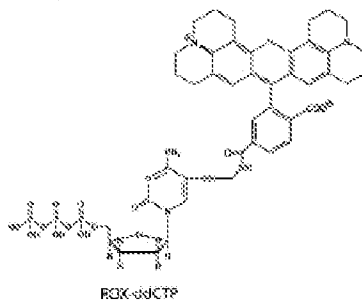


Fig. 2. Nucleotides and nucleotide analogs used for study of Yeast DNA polymerase pre-steady-state kinetic reactions. The structure and name of nucleotide analogs and the chemical structure of nucleotide analogs are shown. The nucleotide analogs are: dCTP, dCTPaS, and CTP (A); nucleotide terminators ddCTP and acyCTP (B); and dye-labeled nucleotide terminators ROX-ddCTP and ROX-acyCTP (C).

substrate or coenzyme concentration, and fitted to the hyperbolic equation:  $v_{obs} = (V_{max}/(K_m + [substrate]))$ , yielding  $V_{max}$ , the maximum rate of nucleotide addition, and  $K_m$ , the dissociation constant for nucleotide binding (58). The activation energy difference between dCTP and nucleotide analog incorporation was calculated by Equation 4 (43):



$$d[\text{dCTP}] = -k_1 [\text{dCTP}] / (K_m + [\text{dCTP}]) \quad (\text{Eq. 1})$$

Single turnover kinetics require saturating enzyme concentrations. We established that 0.10  $\mu\text{M}$  Vent DNA polymerase was sufficient under the reaction conditions described by demonstrating that the rates of dCTP incorporation were the same using Vent DNA polymerase concentrations of 0.10, 0.20, and 0.40  $\mu\text{M}$  (data not shown).

**Measurement of Pyrophosphatase Activity by  $^{32}\text{P}$  Assay.** To measure the rate of DNA elongation by pyrophosphatase, a Vent DNA polymerase (0.10  $\mu\text{M}$ ) was equilibrated with the DNA substrate (0.050  $\mu\text{M}$ ) in 1X Thermofluor buffer and then added with PP<sub>i</sub> in 1X Thermofluor buffer at 66 °C using rapid quench technique as described above. The extent of pyrophosphatase at each time point was calculated by measuring the radioactivity of each DNA species by the number of phosphonate bands hydrolyzed to generate that species.  $k_{\text{cat}}$  and  $K_m$  were derived using fitting protocols analogous to those described above for nucleotide addition.

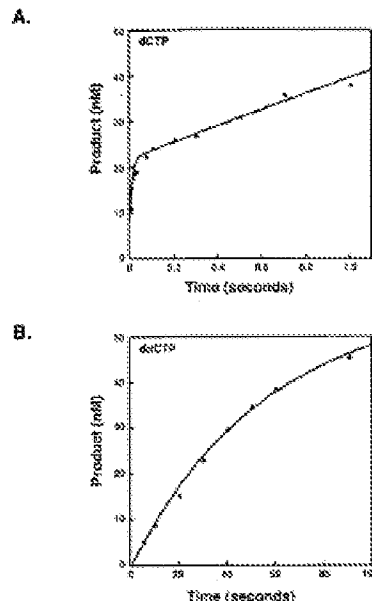
#### RESULTS

**Analysis of dNTP Incorporation by Vent DNA Polymerase.** Previous studies with ExoII DNA polymerases have shown that the steady-state rate-limiting step for addition of a single correctly paired dNTP follows pyrophosphate bond formation (8, 13, 44, 48). Consequently, the first round of polymerization occurs more rapidly than subsequent rounds, resulting in a rapid initial burst of product. Incorporation of dCTP by Vent DNA polymerase displayed a burst pattern similar to those seen with ExoII and AmpliTaq4S DNA polymerases, with a rapid burst ( $k_{\text{burst}} = 60 \text{ s}^{-1}$ ) followed by slow steady-state turnover ( $k_{\text{ss}} = 0.80 \text{ s}^{-1}$ ) (Fig. 3A and Table I). As indicated above, the burst is diagnostic for a rate-limiting step following bond formation, extension; its amplitude is equal to the concentration of active enzyme, indicating that >80% of the Vent DNA polymerase preparation was active. Under similar conditions, Vent DNA polymerase failed to show a significant burst with dGCTP (Fig. 3B) or CTP (data not shown) incorporation. These data suggest that the rate-limiting step during nucleotide analog incorporation lies downstream of dNTP. Upon substitution of dGCTP with dACTP, both Vent and Vent<sup>exo</sup> DNA polymerases showed a 10- and 5-fold (this study) increase in  $k_{\text{burst}}$  and  $k_{\text{ss}}$ , respectively (Table I), consistent with an altered rate-limiting step.

Determinations of  $K_m$  and  $k_{\text{cat}}$  for dCTP addition by Vent DNA polymerase gave kinetic parameters similar to those determined for other DNA polymerases (Fig. 4A and Tables II and III). The relatively high  $K_m$  for nucleotides ( $K_m = 76 \text{ } \mu\text{M}$ ) is similar to the  $K_m$  for nucleotides determined in multiple turnover steady-state measurements ( $K_m = 45 \text{ } \mu\text{M}$ ) (31). Kinetic constants show little dependence on nucleotide identity, as similar Vent DNA polymerase binding ( $K_m = 58 \text{ } \mu\text{M}$ ) and rate ( $k_{\text{cat}} = 0.8 \text{ s}^{-1}$ ) constants were observed for dGTP incorporation. Substitution of dCTP with dCTP/dS had little effect on binding ( $K_m$ ) or pyrophosphate bond formation ( $k_{\text{cat}}$ ); thus, the polymerase displays a minimum kinetic elemental effect ( $k_{\text{cat}}/K_m \text{ dCTP} = 0.80$ ) (Table II).

**Analysis of Vent DNA Polymerase-catalyzed Pyrophosphatase.** To examine Vent DNA polymerase pyrophosphatase activity, we monitored degradation of a  $^{32}\text{P}$ -labeled oligonucleotide duplex in the presence of increasing concentrations of PP<sub>i</sub>. The dependence of the rate of Vent DNA polymerase pyrophosphatase on PP<sub>i</sub> concentration yielded an equilibrium dissociation constant for PP<sub>i</sub> binding of  $K_m = 590 \text{ } \mu\text{M}$  and a maximum velocity of  $k_{\text{cat}} = 2.1 \text{ s}^{-1}$  (Table IV).

**Analysis of Ribonucleotide and Nucleotide Analog Incorporation by Vent DNA Polymerase.** Kinetic parameters of ribonucleotide incorporation were determined to analyze the effect of the presence of a 2'-OH ribonucleotide on polymerization. Vent DNA polymerase discriminated strongly against CTP incorporation over a 10-fold reduced binding affinity ( $K_m = 1.00 \text{ } \mu\text{M}$ )



**Fig. 3. Time-course plots of dCTP and dGCTP incorporation by Vent DNA polymerase.** Incorporation of 200 nM dCTP (A) or dGCTP (B) was monitored as described under "Experimental Procedures." Product (nanomoles) is plotted versus time and fit to the hyperbolic equation:  $y = \frac{a}{1 + e^{-b(x-c)}}$  for A, the Vent DNA polymerase dCTP burst amplitude ( $a$ ) was 21 nM, the burst rate ( $b$ ) was equal to 66  $\text{s}^{-1}$ , and the steady-state rate ( $c$ ) was equal to 0.8  $\text{s}^{-1}$ . In B, the first-order initial rate of dGCTP incorporation was 0.5  $\text{s}^{-1}$ .

and a 20-fold slower rate of nucleotide addition ( $k_{\text{cat}} = 0.160 \text{ s}^{-1}$ ) (Table II). Comparison of CTP and dCTP parameters (expressed as the ratio of catalytic efficiencies:  $(k_{\text{cat}}/K_m)_{\text{CTP}}/(k_{\text{cat}}/K_m)_{\text{dCTP}}$ ) revealed that Vent DNA polymerase preferred dCTP over CTP by 800-fold.

In contrast to CTP, discrimination by Vent DNA polymerase against dGCTP and dACTP was almost exclusively due to a slower rate of nucleotide addition, with  $K_m$  values for dCTP, dGCTP, and dACTP being roughly equal (Fig. 4B and Table II). Indeed, the approximate 30-fold preference for dACTP over dGCTP incorporation can almost entirely be attributed to steps measured by  $k_{\text{cat}}$ .

Similar experiments with Klenow fragment DNA polymerase showed a 33,000-fold higher discrimination against dACTP, affecting steps measured by both  $K_m$  and  $k_{\text{cat}}$ . The Klenow fragment DNA polymerase equilibrium binding constant for dACTP was increased by 20,000 compared with dCTP and dGCTP, whereas  $k_{\text{cat}}$  for dACTP incorporation was no

Enzyme	Starchase (mM)	Asparagine (mM)	Glucose (mM)	Sucrose (mM)
None	50 ± 40	0.50 ± 0.69	0.47 ± 0.35	0.04 ± 0.024
Yeast extract	55 ± 5	3.10 ± 0.61	6.35 ± 1.06	0.04 ± 0.012
Helicobacter	200 ± 40	2.3 ± 0.2	ND	ND
Agaricus bisporus*	50 ± 2	2.5 ± 0.3	ND	ND

**A.**

Y-axis:  $k_{cat}$  ( $s^{-1}$ )

X-axis:  $[dCTP]$  ( $\mu M$ )

**B.**

Y-axis:  $k_{app}$  ( $s^{-1}$ )

X-axis:  $[dCTP]$  ( $\mu M$ )

ROZ-dGTP and RUC-azdTTP incorporation—Previous studies have used RUC derivatized dGTP and azdTTP to measure

08-020164

The single turnover parameters for Pco, DHA polymerase with the normal  $(\text{C}^{15})$  substrates are similar to those of other Family A and B polymerases, both mesophilic and thermophilic, as shown in Table III.  $K_m$  and  $k_{cat}$  values differ by

<sup>c</sup> H. Jiang, M. S. R. Khan, and W. E. Cook, unpublished data.<sup>c</sup> H. Yang, M. M. H. B. Elmaghrabi, and W. E. Jank, unpublished data.

**Table 8**  
*Steady-state kinetic constants for nucleotide analog incorporation by Vent and Vent<sup>mut</sup> DNA polymerases*  
In almost all cases, the kinetic parameters for Vent and Vent<sup>mut</sup> DNA polymerases are from at least two independent determinations except where indicated by footnote b and are reported as the mean  $\pm$  S.D.

Nucleotide	Vent DNA polymerase				Vent <sup>mut</sup> DNA polymerase			
	$k_p$ $\mu\text{M}^{-1}\text{s}^{-1}$	$k_{pp}$ $\text{s}^{-1}$	$k_{pp}/k_p$ $\mu\text{M}$	Discrimination <sup>a</sup>	$k_p$ $\mu\text{M}^{-1}\text{s}^{-1}$	$k_{pp}$ $\text{s}^{-1}$	$k_{pp}/k_p$ $\mu\text{M}$	Discrimination <sup>a</sup>
dCTP	$76 \pm 7$	$66 \pm 3$	$0.8 \times 10^3$		$77 \pm 9$	$58 \pm 3$	$1.3 \times 10^3$	
dCTPdG	$102 \pm 45$	$82 \pm 13$	$0.8 \times 10^3$	1.4	$88 \pm 12$	$58.0 \pm 0.9$	$6.1 \times 10^2$	1.9
CTP	$1366 \pm 100$	$0.189 \pm 0.001$	$1.5 \times 10^3$	8940	$988 \pm 92$	$0.75 \pm 0.02$	$1.6 \times 10^3$	450
dATP	$42 \pm 7$	$0.18 \pm 0.01$	$4.5 \times 10^3$	230	$35 \pm 4$	$0.92 \pm 0.02$	$1.6 \times 10^4$	40
anyCTP	$85 \pm 20$	$7.6 \pm 1.1$	$0.7 \times 10^3$	30	$28.0 \pm 0.4$	$1.0 \pm 0$	$5.4 \times 10^3$	1.4
Klenow-dCTP	$30 \pm 2$	$0.009 \pm 0.003$	$3 \times 10^3$	320	$8.6^b$	$0.2^b$	$2.6 \times 10^3$	30
Klenow-anyCTP	$8.6 \pm 2.9$	$0.04 \pm 0.01$	$2.5 \times 10^3$	4	$0^b$	$0^b$	$1.8 \times 10^3$	4.5

<sup>a</sup> Selectivity between incorporation of dCTP and other dNTPs is the ratio of the efficiency of dCTP incorporation ( $k_{pp}/k_p$ ) to the efficiency of CTP, dATP, or anyCTP incorporation.

<sup>b</sup> The kinetic parameters for Vent<sup>mut</sup> DNA polymerase are from single determinations.

**Table 9**  
*Steady-state kinetic constants for nucleotide analog incorporation by DNA polymerases*  
The kinetic parameters for Vent and Vent<sup>mut</sup> DNA polymerases are from at least two independent determinations and are reported as the mean  $\pm$  S.D. SD, not determined.

Nucleoside	dCTP				CTP				dATP				anyCTP			
	$k_p$ $\mu\text{M}^{-1}\text{s}^{-1}$	$k_{pp}$ $\text{s}^{-1}$	$k_{pp}/k_p$ $\mu\text{M}$	Discrimination <sup>a</sup>	$k_p$ $\mu\text{M}^{-1}\text{s}^{-1}$	$k_{pp}$ $\text{s}^{-1}$	$k_{pp}/k_p$ $\mu\text{M}$	Discrimination <sup>a</sup>	$k_p$ $\mu\text{M}^{-1}\text{s}^{-1}$	$k_{pp}$ $\text{s}^{-1}$	$k_{pp}/k_p$ $\mu\text{M}$	Discrimination <sup>a</sup>	$k_p$ $\mu\text{M}^{-1}\text{s}^{-1}$	$k_{pp}$ $\text{s}^{-1}$	$k_{pp}/k_p$ $\mu\text{M}$	Discrimination <sup>a</sup>
Vent	$76 \pm 7$	$66 \pm 3$	$1.00 \pm 0.00$	8700	$48 \pm 7$	$0.31 \pm 0.01$	$650$	$81 \pm 15$	$7.6 \pm 1.1$	$0.07 \pm 0.01$	$1.0 \times 10^3$	30				
Vent <sup>mut</sup>	$77 \pm 9$	$58 \pm 3$	$0.75 \pm 0.02$	430	$18 \pm 3$	$0.20 \pm 0.02$	$11$	$38.0 \pm 0.4$	$1.0 \pm 0$	$0.01 \pm 0.01$	$1.0 \times 10^3$	1.4				
Klenow	$88 \pm 12$	$58.0 \pm 0.9$	$0.65 \pm 0.02$	6200	$4200 \pm 900$	$0.17 \pm 0.02$	$21,000$	$21,000$	$8.6^b$	$0.2^b$	$2.6 \times 10^3$	30				
Klenow	$30 \pm 2$	$0.009 \pm 0.003$	$3 \times 10^3$	320	$8.6^b$	$0.2^b$	$2.6 \times 10^3$	30								
Klenow	$8.6 \pm 2.9$	$0.04 \pm 0.01$	$2.5 \times 10^3$	4												

<sup>a</sup> Selectivity between incorporation of dCTP and other dNTPs is the ratio of the efficiency of dCTP incorporation ( $k_{pp}/k_p$ ) to the efficiency of CTP, dATP, or anyCTP incorporation.

<sup>b</sup> Ref. 20.

<sup>c</sup> Ref. 20.

<sup>d</sup> Ref. 27.

<sup>e</sup> Ref. 28.

**Table 10**  
*Reverse-primase activity*  
The kinetic parameters for Vent and Vent<sup>mut</sup> DNA polymerases are from at least two independent determinations and are reported as the mean  $\pm$  S.D.

Nucleoside	$k_{pp}$		$k_{pp}/k_p$	
	$\mu\text{M}^{-1}\text{s}^{-1}$	$\text{s}^{-1}$	$\mu\text{M}$	$\mu\text{M}^{-1}\text{s}^{-1}$
Vent	$360 \pm 168$	$1.10 \pm 0.37$	$3.2 \times 10^3$	388
Vent <sup>mut</sup>	$1760 \pm 400$	$0.83 \pm 0.21$	$2.7 \times 10^3$	1300
Klenow <sup>a</sup>	$26,000$	$0.33$	$1.3 \times 10^3$	230,000
Klenow <sup>b</sup>	$280$	$0.23$	$1.3 \times 10^3$	6600

<sup>a</sup> A comparison of DNA polymerase activity with pyrophosphatase (pylase) is given by the ratio of DNA polymerase efficiency ( $k_{pp}/k_p$ ) to  $k_{pp}/k_p$  divided by the efficiency of pyrophosphatase ( $k_{pp}/k_p$ ).

<sup>b</sup> Ref. 28.

<sup>c</sup> Ref. 28.

<10-fold for all polymerases tested, with no clear division between Family A and B DNA polymerases. Furthermore, the Family A Klenow fragment and Family B Vent and RB69 DNA polymerases carry out the reverse reaction of DNA polymerization, pyrophosphorylation, with similar rates ( $k_{pp}$ ), and Klenow fragment and Vent DNA polymerases share comparable  $k_p$  binding constants ( $k_{pp}/k_p$ ) (Table IV). Similarities in nucleotide incorporation kinetics and active site structure underscore the evolution of DNA polymerases to efficiently carry out DNA replication and repair. Significant kinetic differences between the polymerases become apparent only when examining nucleotide analog incorporation and extended effects, as detailed below.

**Rejection of dNTPs.**—Despite a similar level of selectivity against dNTPs, their discrimination is supplied almost exclusively by differences associated with  $k_{pp}$  for Klenow fragment DNA polymerase, whereas Vent DNA polymerase shows not only  $k_{pp}$  effects, but also a 10-fold weaker ground state binding of the

nucleotide. RB69 DNA polymerase also shows effects in both  $k_p$  and  $k_{pp}$ , achieving an even higher discrimination by virtue of a 200-fold weaker ground state binding. Discrimination against dNTPs, in large part, been attributed to a steric clash between the 3'-OH and a conserved amino chain in the polymerase active site (28, 30, 33). The kinetic parameters suggest that the steric clash is first encountered in the ground state nucleotide binding by Vent and RB69 DNA polymerases, but does not affect Klenow fragment DNA polymerase until the transition state of the reaction. This could occur, for example, if the  $k_p$  term for Klenow fragment DNA polymerase primarily measures binding prior to a conformational shift that engages the 3'-OH bearing machinery.

**Difformity of dNTPs.**—When incorporating ddCTP, RB69 DNA polymerase discriminates at the level of both  $k_p$  and  $k_{pp}$ . Discrimination by Vent, Klenow fragment, and Klenow fragment DNA polymerase with a 206-amino acid B-terminal deletion (B206) DNA polymerase is almost exclusively in the

steps measured by  $k_{pol}$  and not those involved in  $K_D$  with Vent DNA polymerase showing less discrimination than the other two polymerases. This parallel behavior appears to reflect a partial lack of 3'-OH involvement in ground state substrate binding rather than a conserved set of nucleotide contacts.

On the surface, the similarity in  $k_{pol}$  values for dNTP incorporation by Vent and Klenow fragment DNA polymerase (32) suggests similar discriminatory mechanisms for these two enzymes, a conclusion reinforced by the absence of an isoelectronic effect with dNTPs as being either enzymes. The simplest interpretation of the lack of an isoelectronic effect with  $\alpha$ -thio-substituted dNTPs with Klenow fragment and Vent DNA polymerase is that a non-classical step(s) preceding phosphodiester bond formation is rate-limiting. Similarly, the lack of a significant isoelectronic effect for Klenow fragment DNA polymerase incorporation of ddNTPs (32) argues that steps preceding phosphodiester bond formation continue to be rate-limiting for that enzyme. In contrast, the isoelectronic effect noted for Vent and Vent<sup>98H</sup> DNA polymerase incorporation of ddNTPs is an indication that the chemistry of phosphodiester bond formation significantly influences the rate-limiting step for these polymerases.

The  $k_{pol}$  rates with both polymerases were significantly slower for ddNTPs than for dNTPs: K480 and 412-fold for Klenow fragment and Vent DNA polymerase, respectively. In the case of Vent DNA polymerase, this must reflect at least a slowing of the isoelectronic rate, whereas for Klenow fragment DNA polymerase, at least the role of non-classical step(s) must be shown. Thus, the pre-thioester rate for Vent DNA polymerase ddNTP incorporation is at least 10-fold faster than comparable steps for Klenow fragment DNA polymerase.

Conserved amino acids positioned within active site Family A or B DNA polymerase active sites probe for correctly base-paired substrates not immediately align phosphates into a geometry required for phosphoryl transfer. As observed by Franklin *et al.* (33) and Yang *et al.* (25) in the *PolB* DNA polymerase ternary crystal structure (and by analogy, in the Vent DNA polymerase active site) (Fig. 1), the dNTP deoxyribose moiety assumes a favorable 3'-endo sugar conformation. This conformation is constrained by hydrogen bonds between the 3'-OH and a main chain amide corresponding to Vent DNA polymerase position 412N and a non-bridging  $\beta$ -phosphate oxygen (Fig. 5, A and B). Nucleotide  $\alpha$ -,  $\beta$ -, and  $\gamma$ -phosphates are further stabilized by direct or water-mediated hydrogen bonds with active site residues (Fig. 5, A and B). The absence of the 3'-OH on ddNTPs disrupts hydrogen bonding with the  $\beta$ -phosphate (and main chain amide), potentially increasing the activation energy required to orient the  $\alpha$ -phosphate for phosphoryl transfer (Fig. 5C). Indeed, the measured energetic difference between dNTP and ddNTP incorporation (18 kcal mol<sup>-1</sup>) is equivalent to that expected for the loss of at least two hydrogen bonds in the dNTP transition state (47).

Although the active site binding network differs, in the Family A Klenow fragment DNA polymerase the dNTP 3'-OH contributes 21 kcal mol<sup>-1</sup> in transition state stabilization, accounting for inefficient ddNTP incorporation (27). This energy loss is counteracted in the closely related  $\Phi$  DNA polymerase active site by a hydroxyl group at Tyr<sup>99</sup> (Klenow fragment DNA polymerase has Phe in the analogous position) that facilitates a hydrogen bond to stabilize the ddNTP  $\beta$ -phosphate in the transition state, re-establishing a hydrogen bonding network similar to interactions formed by dNTP (12). As a result,  $\Phi$  DNA polymerase selectivity between dNTP and ddNTP is greatly reduced, as is the selectivity of the analogous  $\Phi$ -type-Tye mutation in both Klenow fragment and  $\Phi$  DNA polymerases (49).

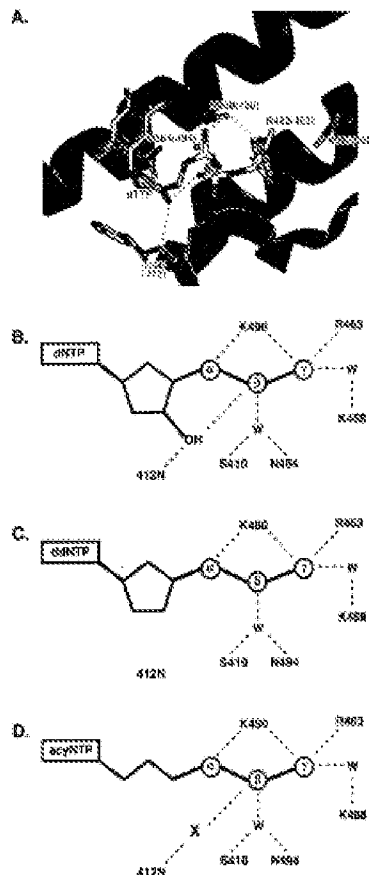


Fig. 5. Active site models of dNTP, ddNTP, and acpNTP incorporation. A, the Klenow fragment DNA polymerase ternary crystal structure shows active site interactions that stabilize the substrate dNTP (33). Vent DNA polymerase incorporation is shown in parentheses. B, a schematic of the Vent DNA polymerase active site interactions that stabilize the dNTP transition state is presented in a perspective view. W, water molecule. C, a model of the Vent DNA polymerase active site with ddNTP (based on loss of hydrogen bonding with the Tyr<sup>98H</sup> main chain amide, dNTP 3'-OH, and non-bridging  $\beta$ -phosphate). D, a model for the binding of acpNTP suggests that, in the absence of a ribose ring, a nonbridging  $\beta$ -phosphate could be a major nucleoside-like, nonbridging hydrogen bonding between the Tyr<sup>98H</sup> main chain amide and non-bridging  $\beta$ -phosphate.

**Nucleoside Analog Incorporation.**—Discrepancy between polymerases is also noted upon acyNTP addition, again suggesting divergent mechanisms for nucleotide recognition and incorporation between the polymerases. Similar to ddNTPs, acyNTPs lack the 3' OH required to establish a hydrogen bonding network between the main chain amide of T<sub>yr</sub><sup>117</sup> and 5'-phosphate of the substrate (Fig. 5D). Kinetic fragment DNA polymerase displays a strong discrimination in both  $K_d$  and  $k_{pol}$  terms, resulting in a selectivity value of 32,800. In this case, the efficiency of acyNTP incorporation is nearly as low as for an incorrect base pair ( $k_{pol}/K_d \sim 240$  and  $96 \text{ s}^{-1} \text{ s}^{-1}$ , respectively) (Table III). A strong bias against acyNTP incorporation has also been noted for Tag DNA polymerase (33).

In contrast, acyNTPs are incorporated by hyperthermophilic archaeal DNA polymerases with only 10-fold lower efficiency than dNTPs. By analogy with ddNTP incorporation by T7 DNA polymerase, it seems reasonable that the space normally occupied by the capor 3' and 3' carbons and associated substituents would be accessible to water molecules, metals, or protein side chains that might establish interactions to compensate for those disrupted by the missing 3'-OH group. The difference in activation free energy between ddNTP and acyNTP incorporation ( $\Delta\Delta G^\ddagger = \Delta G^\ddagger_{acyNTP} - \Delta G^\ddagger_{ddNTP} = 15 \text{ kJ mol}^{-1}$ ) is equivalent to a gain of two additional hydrogen bonds, which could be provided by hydrogen bonding between the main chain 432 amide, a protein water, and an acyNTP 5'-phosphate coordinating oxygen to metal interactions that exist in the dNTP active site (Fig. 8B). As the space here, we cannot rule out stabilizing interactions arising from residues near the active site normally excluded by the ribose 2' and 3' carbons that are absent in acyNTP. Clearly, three-dimensional structural analysis will be necessary for a full understanding of the structural changes important for Yeast DNA polymerase incorporation of acyNTPs.

**Dye-Substituted Nucleotides.**—Dye-substituted nucleotides have been useful in a variety of enzymological applications (40, 43). Not surprisingly, given the diversity of dye structures and charges, dye-substituted nucleotides are accepted by DNA polymerases with varying efficiencies (28, 29, 33, 36). Previous studies identified nucleotide derivatives containing the fluorescent dye BOD as being more efficiently incorporated by Yeast DNA polymerase than the parental nucleotides lacking the dye (28, 34). In the current kinetic studies, the magnitude of enhanced dye-substituted nucleotide addition was much lower than previously estimated in semiquantitative gel denaturation assays, even though those same gel titration assays gave good agreement with the relative incorporation efficiency of ddNTP and acyNTP substrates (33). BOD substitution of the nucleotide results in a 5–10-fold lower  $K_d$ , suggesting that contacts in or adjacent to the Yeast DNA polymerase active site stabilize dye binding. However, at the same time,  $k_{pol}$  is reduced, suggesting that one consequence of the enhanced binding is to slow nucleotide addition. Thus, substrate incorporation is a balance of both binding and catalysis: a substrate bound with too high affinity requires higher activation energy for efficient turnover by the polymerase.

**Yeast DNA Polymerase Pre-steady-state Kinetics.**—Previous studies identified a Yeast DNA polymerase variant Y4488L with enhanced nucleotide analog incorporation properties (33, 34). Efficient dNTP incorporation by Y4488L DNA polymerase is characterized by higher binding affinity ( $K_d$ ), nucleotide transfer rate ( $k_{pol}$ ), and rate-limiting step compared with Yeast DNA polymerase, presumably reflecting the conservation of residues actively involved in coordinating the incoming dNTP. In contrast, each of the nucleotide analogs tested with Y4488L DNA polymerase have higher binding affinity and

factor rates of phosphoryl transfer than the unmodified polymerase. Energy differences between Yeast and Y4488L DNA polymerase incorporation of ddNTP or acyNTP are modest ( $\Delta\Delta G^\ddagger_{ddNTP} \sim 4.5 \text{ kJ mol}^{-1}$  and  $\Delta\Delta G^\ddagger_{acyNTP} \sim 4.8 \text{ kJ mol}^{-1}$ ), suggesting that subtle hydrophobic or hydrogen bond-mediated effects could account for enhanced analog incorporation.

One hypothesis to account for these effects associates the Y4488L variant as being closer to the relaxed conformation, thus facilitating incorporation of analogs. The residue analogous to Y4488 in the BOD DNA polymerase crystal structure points away from the active site and lies at the interface between the sides of the polymerase and as a helix that must make a 60° rotation to form the closed complex (Fig. 1). In Yeast DNA polymerase, positioning a longer leucine residue at the position normally occupied by alanine in the Y4488L may shift equilibrium from the open toward the closed conformation, thus reducing the activation energy for both binding and nucleotide transfer. This comes at a price: direct experimental demonstration that subsequent turnover by the Y4488L variant is inhibited, perhaps reflecting hindrance of the transition from closed to open states required for release and/or binding of the template and dNTP. This proposal does not, however, easily account for the fact that pre-steady-state kinetics for the natural substrates are unaltered in this variant. Alternatively, resolution of this discrepancy may lie in the greater ability of this variant to overcome distortions in the nucleotide-binding site, distortions that are not present when the correct nucleotide is bound.

In summary, from these comparative studies, we observed that kinetics of ddNTP incorporation pathways are conserved among Family A and B DNA polymerases despite diversity in primary amino acid sequence, fluorocoding, stability, and biological roles. However, differences in acyNTP and other nucleotide analog catalytic efficiencies in Kinase fragment, Yeast, and other DNA polymerases illuminate fundamental differences underlying the kinetic pathways for DNA polymerization. As more DNA polymerases are studied kinetically, it is apparent that subtle structural variations in the active site influence how nucleotides are bound and positioned for catalysis.

**Acknowledgments.**—We are grateful to Charles Richardson for providing this research as part of the National Institute of Health Program of Molecular Biophysics (to A. P.). We also thank Paul Smith for providing the labeled compounds and for helpful discussions, Peter Henshaw (Applied Biosystems) for useful discussions and technical advice, Nicola Nithila, Brianne Pruthi, and Tom Evans for critical review of this manuscript, and Chris Brown, Lucia Greenough, and John Brown for providing expert technical assistance. We are also indebted to the Center for Polymerase's supportive research environment at New England Biolabs Inc.

# REFERENCES

1. Kornberg, A. (1980) *DNA Replication*, pp. 57–97, W. H. Freeman & Co., San Francisco.
2. Joyce, D. M., and Smith, T. A. (1987) *Annu. Rev. Biochem.* 56, 717–749.
3. Smith, T. A. (1990) *J. Biol. Chem.* 265, 17333–17339.
4. Kornberg, D. R., and Smith, T. A. (1991) *Science* 251, 955–957.
5. Fink, J., Fink, J. P., Smith, T. A., and Kornberg, D. R. (1991) *J. Biol. Chem.* 266, 7617–7621.
6. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 14281–14285.
7. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1990) *J. Biol. Chem.* 265, 14281–14285.
8. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
9. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
10. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
11. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
12. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
13. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
14. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
15. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
16. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
17. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
18. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
19. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
20. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
21. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
22. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
23. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
24. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
25. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
26. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
27. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
28. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
29. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
30. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
31. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
32. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
33. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.
34. Kornberg, D. R., Smith, T. A., Fink, J. P., and Kornberg, D. R. (1991) *Science* 251, 955–957.

- 507-512.
16. Song, W., Tapscott, W. C., and Wang, T. S. N. (1999) *J. Biol. Chem.* 274, 29781-29787.
17. Sze, M. W., Menad, N. G., Cooper, N. J., and Greenleaf E. J. (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 3378-3383.
18. Hughes, M. P., Greenleaf, A., Dyer, S. A., Leary, P., Amherst-Davis, W., Hynes, S., and Cooper, S. (1999) *Proc. Natl. Acad. Sci. U. S. A.* 96, 5991-5995.
19. Zhou, S., deGuzman, L., Marmè, D., Laghetti, L., Lachon, B., and Klenow, J. (1994) *Nucleic Acids Res.* 22, 1180-1185.
20. Searle, A., Lamm, A. M., Sherr, L., and Biles, M. (1999) *J. Biol. Chem.* 274, 2017-2021.
21. Sheng, Y., Huang, L., and Biles, M. (1999) *J. Biol. Chem.* 274, 45-48.
22. Rodriguez, A. C., Park, M. W., Park, T., and Hwang, L. S. (2000) *J. Biol. Chem.* 275, 407-409.
23. Fiedler, M. C., Wang, J., and Smith, D. A. (2001) *Cell* 105, 291-297.
24. Rodriguez, R., Rodriguez, M., Rodriguez, S., Zhang, W., Sanchez, V., Lopez, T., and Rio, E. (2001) *J. Biol. Chem.* 276, 401-407.
25. Yang, G., Fiedler, M., Li, J., Gu, T. C., and Klenow, J. (1992) *Nucleic Acids Res.* 20, 5925-5930.
26. Aggeler, M., Og, S., Grollman, A. P., and Joyce, G. M. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 3552-3557.
27. Alonso, M., Fiedler, M. C., and Joyce, G. M. (1999) *J. Biol. Chem.* 274, 349-355.
28. Brown, J. W. (1999) *Nucleic Acids Res.* 27, 1512-1519.
29. Hori, S. E., and Rhee, P. B. (1997) *Environ. J.* 12, 619-624.
30. Yang, G., Fiedler, M., Li, J., Gu, T. C., and Klenow, J. (1992) *Nucleic Acids Res.* 20, 1079-1083.
31. Wang, W., Klenow, J. S., and Loh, M. E. (1992) *J. Biol. Chem.* 267, 1060-1065.
32. Gordon, A. F., and Jahn, W. E. (1999) *Nucleic Acids Res.* 27, 3545-3557.
33. Gordon, A. F., and Jahn, W. E. (1999) *Nucleic Acids Res.* 27, 3558-3570.
34. Rhee, S. J., Yang, M. J., Klenow, J. S., Park, J. H., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3571-3583.
35. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3584-3596.
36. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3597-3609.
37. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3610-3622.
38. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3623-3635.
39. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3636-3648.
40. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3649-3661.
41. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3662-3674.
42. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3675-3687.
43. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3688-3700.
44. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3701-3713.
45. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3714-3726.
46. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3727-3739.
47. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3740-3752.
48. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3753-3765.
49. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3766-3778.
50. Fiedler, M., Rhee, S. J., and Klenow, J. S. (1999) *Nucleic Acids Res.* 27, 3779-3791.

## Crystal Structure of DNA Polymerase from Hyperthermophilic Archaeon *Pyrococcus kodakaraensis* KOD1

Hiroshi Hashimoto<sup>1</sup>, Motomu Nishioka<sup>2</sup>, Shinsuke Fujiwara<sup>3</sup>  
Masahiro Takagi<sup>2</sup>, Tadayuki Imanaka<sup>2</sup>, Tsuyoshi Inoue<sup>1</sup>  
and Yasushi Kai<sup>1\*</sup>

<sup>1</sup>Department of Materials Chemistry and

<sup>2</sup>Department of Biotechnology Graduate School of Engineering, Osaka University 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

<sup>3</sup>Department of Synthetic Chemistry and Biological Engineering, Kyoto University Nishikyo-machi Sakyo-ku Kyoto 606-8501, Japan

The crystal structure of family B DNA polymerase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 (KOD DNA polymerase) was determined. KOD DNA polymerase exhibits the highest known extension rate, processivity and fidelity. We carried out the structural analysis of KOD DNA polymerase in order to clarify the mechanism of these enzymatic features. Structural comparison of DNA polymerases from hyperthermophilic archaea highlighted the conformational differences to *Thermost* domains. The thumb domain of KOD DNA polymerase shows an "open" conformation. The fingers subdomain possessed many basic residues at the side of the polymerase active site. The residues are considered to be accessible to the incoming dNTP by electrostatic interaction. A 3-hairpin motif (residues 243–245) consists from the backbone (B<sub>1</sub>) domain as seen in the editing complex of the KOD DNA polymerase from *Thermost* (KOD). Many arginine residues are located at the forked-point (the junction of the template-binding and editing cleft) of KOD DNA polymerase, suggesting that the basic environment is suitable for partitioning of the primer and template DNA duplex and for stabilizing the partially melted DNA structure in the high-temperature environments. The stabilization of the melted DNA structure at the forked-point may be correlated with the high PCR performance of KOD DNA polymerase, which is due to low error rate, high elongation rate and processivity.

© 2003 Academic Press

**Keywords:** archaea; crystal structure; family B DNA polymerase; "forked-point"; KOD DNA polymerase

\*Corresponding author

### Introduction

DNA polymerases are a group of enzymes that use single-stranded DNA as a template for the synthesis of the complementary DNA strand. These enzymes are multifunctional, with both synthetic (polymerase) and one or two degradative modes (5'–3' and/or 3'–5' exonucleases) and play an essential role in nucleic acid metabolism including the processes of DNA replication, repair and recombination. Many DNA polymerase genes have been cloned and sequenced. Amino acid sequences deduced from their nucleotide sequences can be classified into four major types: *E. coli* type

DNA polymerase I (family A), *E. coli* DNA polymerase II (family B), *E. coli* DNA polymerase III (family C) and others (family X).<sup>1</sup> Recently, a new family of DNA polymerases has been identified, all members of this family contain five highly conserved motifs, I–V, and several of these polymerases participate in lesion bypass.<sup>2</sup> This family is called the *Unid./Dex* family.<sup>3</sup> Family B DNA polymerases include eukaryotic DNA polymerase  $\alpha$ ,  $\delta$ , and  $\epsilon$ , which are thought to be components of the replisome and to carry out chromosomal DNA replication. Archaeal proteins involved in gene expression, such as those for DNA replication, transcription, and translation, have been found to be similar to those from eukarya. Therefore, the archaeal system of gene expression is a simplified model of the eukaryotic system. In contrast, the

\*E-mail address of the corresponding author: [kai@chem.sci.osaka-u.ac.jp](mailto:kai@chem.sci.osaka-u.ac.jp)

cellular appearance and organization of archaea are more like those of bacteria.

The first crystal structure of a family B DNA polymerase to be obtained was that of bacteriophage KB9 DNA polymerase (KB9 DNA polymerase).<sup>4</sup> The first crystal structure of archaeal DNA polymerase was TNA DNA polymerase from *Thermococcus gorgonius* (Tgo DNA polymerase).<sup>5</sup> The editing complex of KB9 DNA polymerase has been reported.<sup>6</sup> Two further crystal structures of archaeal family B DNA polymerases have recently been reported: Tso DNA polymerase from *Ditella furiosus* sp. Tsk<sup>7</sup> and 9°N-7 DNA polymerase from *Thermococcus* sp. 9°N-7.<sup>8</sup>

The Pyrococcus furiosus KOD1 is a hyperthermophilic archaeon, with an optimum growth temperature of 95°C.<sup>9</sup> Enzymes produced in KOD1 were reported to be extremely thermostable and to have eukaryotic characteristics.<sup>9</sup> The optimum temperature of KOD DNA polymerase is 75°C, similar to that of DNA polymerase obtained from *Pyrococcus furiosus* (Pfu DNA polymerase). KOD DNA polymerase, however, exhibits the higher extension rate (100–130 nucleotides/second) and processivity (>300 bases) five times and ten to 15 times higher than those of Pfu DNA polymerase, respectively.<sup>10</sup> Thermostable DNA polymerases are expected to be suitable reagents for Polymerase Chain Reaction (PCR). KOD DNA polymerase is, therefore, suitable for DNA amplification by such means. Indeed, KOD DNA polymerase is widely used in rapid and accurate PCR systems (TOYOBO Ltd., Japan).

Although structures of three archaeal DNA polymerases have been determined as described above, no structural information relating to elongation rate, processivity or fidelity is provided. We carried out the structural analysis of KOD DNA polymerase in order to clarify the mechanism of enzymatic features of KOD DNA polymerase, which are the highest extension rate, processivity and fidelity. Here, we report the crystal structure of DNA polymerase from the hyperthermophilic archaeon *Pyrococcus kodakarensis* KOD1. The three-dimensional structure of this KOD DNA polymerase may provide useful information to clarify the mechanisms for rapid and accurate reaction. In addition, this information may contribute to the improvement of the PCR properties of enzymes already in use such as thermostability, error rate, elongation rate and processivity, or for designing new enzymes for PCR as well as DNA replication by family B DNA polymerases.

## Results and Discussion

### Overall structure

KOD DNA polymerase has a doli-like shape with dimensions 68 Å × 80 Å × 100 Å and is made up of distinct domains and subdomains: N-terminal (N-ter: 1–188, 327–368, violet), Exonuclease (Exo: 131–336, blue), Polymerase (Pol)

domain including the Palm and Fingers subdomains (369–449, 300–387, brown; and 420–499, green, respectively) and the Thumb domain including the (thumb-1) and (thumb-2) subdomains (588–773, red) (Figure 1(a)). The polymerase active site, including three conserved carboxylates, (Asp426, Asp460 and Asp542) is located in an anti-parallel β-sheet in the Palm subdomain. The exonuclease active site contains two conserved carboxylates (Asp141 and Glu143) and is located in an anti-parallel β-sheet in the Exo domain. The Polymerase and exonuclease active sites on the molecular surface are indicated by P and E, respectively (see Figure 4). Structural comparisons of archaeal DNA polymerases (KOD, Tgo and 9°N-7 DNA polymerases) are shown in Figure 1(b). The structural architectures of the proteins are identical, but the orientation of the domains and subdomains is different. In the case of the KOD DNA polymerase (red), the Thumb domain is shifted to make an "open" conformation and the position of the Palm domain neighboring the root of the Thumb domain is slightly shifted as a result of the large movement of the Thumb domain in comparison to other archaeal DNA polymerases. Table 1 shows the averaged temperature factors of the domains and subdomains in the crystal structure of KOD DNA polymerase. The value of the Thumb domain was markedly higher than the others. The structures of many residues in the Thumb-2 subdomain are not defined, because the orientation of the subdomain is highly disordered. Therefore, it is thought that the structure of KOD DNA polymerase described here provides information for the DNA-trait, most relaxed conformation. The structure of the editing complex of KB9 DNA polymerase revealed that newly synthesized duplex DNA is grasped by the Pol and Thumb domains. Although the orientation of the Thumb domain is potentially highly flexible, the orientation may be fixed when it binds to the primer-template duplex.

### Polymerase domain

The Pol domain is made up of the Fingers and Palm subdomains and has an "L-like" shape (Figure 2(a)). The polymerization mechanism has been studied mainly on family A DNA polymerases (220–3). A structural basis for a metal-

Table 1. Averaged temperature factors

Domain	Temperature factor (Å <sup>2</sup> )
N-ter	38.1
Exo	32.7
Pol	
Fingers	49.3
Palm	32.8
Thumb	62.7
Overall	39.9



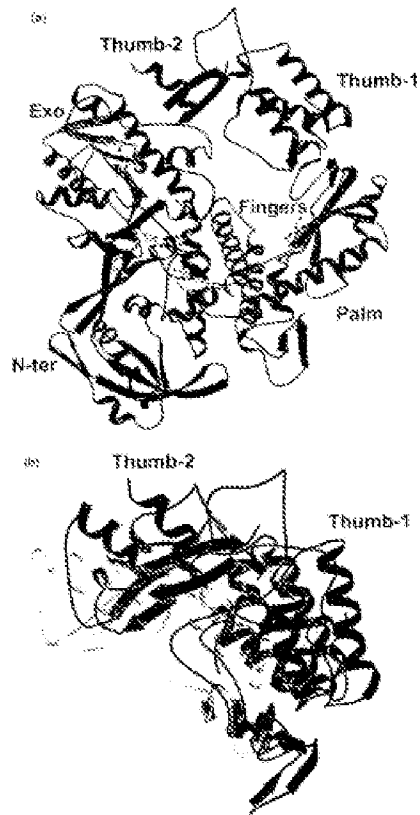


Figure 1. (a) Ribbon structure of KOD DNA polymerase. The structure is composed of domains and subdomains, which are  $\beta$ -sheeted (N-ter, violet),  $\alpha$ -helical (Exo, blue),  $\alpha$ -helical (Palm, pink), Polymerase (Palm) domain including the Palm domain; and Fingers (green) subdomain and the Thumb domain (red), including the Thumb-1 and Thumb-2 subdomains. Conserved carboxylate residues in Polymerase and Exonuclease active site are shown by ball-and-stick models. (b) Conformational superposition of Thumb domains among three bacterial DNA polymerases: Red, KOD DNA polymerase; blue, Taq DNA polymerase; and green, Pfu DNA polymerase. The comparison shows that the Thumb domain of KOD DNA polymerase displays the most "rigid" conformation.

proposed mechanism of phosphate transfer was provided by the bacteriophage T7 DNA replication complex.<sup>11</sup> The complex structure shows that two metal ions are bound by strictly conserved aspartates (Asp473 and Asp553, which correspond to Asp463 and Asp542 in KOD DNA polymerase)

extended from the anti-parallel  $\beta$ -sheet of the Palm domain. The phosphate group of incoming dGTP is held by the metal ions and the four basic residues extending from the Fingers subdomain (His505, Arg518 and Lys522). The crystal structure of two binary complexes of the large fragment of

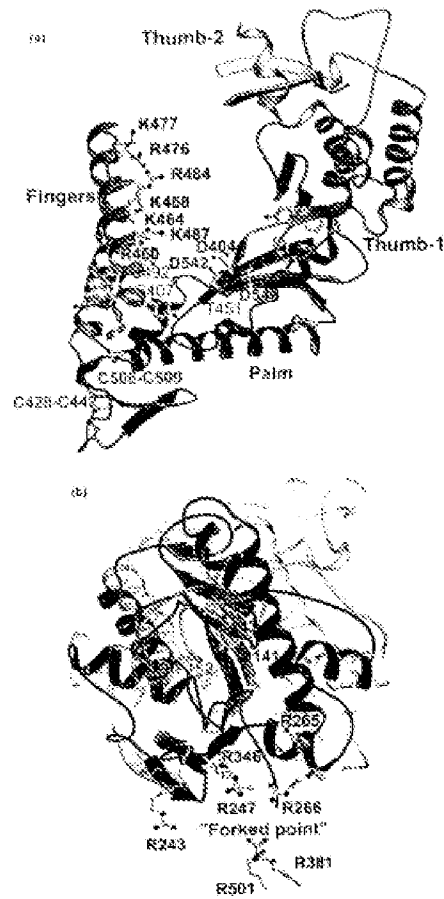


Figure 2. (a) Ribbon representation of the Pol domain. The domain is made up of Fingers and Thumb subdomains. Conserved carboxylate residues (D404, D542 and D547) are represented by ball-and-stick models. Basic residues are represented by ball-and-stick models, which stand in a line in the Fingers subdomain, facing the polynucleotide active site. K464 are replaced by alanine, because of the ambiguity of its electron density. Two disulfide bonds are displayed (C428-C442 and C508-C509). Aromatic residues adjacent to a glycine residue, represented by ball-and-stick models, are localized in the groove of the subdomain. The Thumb domain is represented by the corresponding model. The C $\alpha$  atoms of the polynucleotide residues, 5607 (for Pol-1), 5492 (for Pol-2) and 1541 (for Pol-3), are represented by violet spheres. (b) Surface representation of KOD DNA polymerase and KOD DNA polymerase (nontransparent model). Conserved carboxylate (D404 and D542) and arginine residues (R473, R477, R545, R266, R243, R381 and R501) in the forked-point of KOD DNA polymerase are represented by ball-and-stick models. The red domain are (disulfide bond) parts forming template-binding and editing clefts. The loop containing His152 is shown in orange. F123 of KOD DNA polymerase and F152 of KOD DNA polymerase are represented by nontransparent and opaque ball-and-stick models, respectively.

*Thermus aquaticus* DNA polymerase I (Klenow); with a primer-template DNA and dNCTP have been reported.<sup>12</sup> The ternary complexes suggest that basic residues of the fingers subdomain fold the phosphate group of the incoming dNTP and the domain induces a conformational change to deliver the incoming nucleotide to the active site. In the case of family B DNA polymerases, the fingers subdomain is composed mainly of one long helix and does not have a joint first appeared in the structures of family A DNA polymerases. Therefore, it seems that in the case of archaeal DNA polymerases, the movement of the Pol domain to deliver dNTP to the active site differs from that of family A DNA polymerases. Kinetic study of K869 DNA polymerase mutants revealed that four residues (Arg482, Lys486, Lys560 and Asn561) of the fingers subdomain affected dNTP incorporation.<sup>13</sup> The residues are conserved in family B DNA polymerases, and correspond to Arg461, Lys464, Lys487 and Asn491 in KOD DNA polymerase, respectively. Furthermore, Lys468, Arg476, Lys477 and Arg484 are located at the tip of the fingers subdomain on the side of the polymerase active site in KOD DNA polymerase (Figure 2a). It is expected that the "gate" of basic residues captures the incoming dNTP, then the dNTP is delivered toward the polymerase active-site pocket by accompanying the movement of the polymerase domain. Two disulfide bonds exist in the connection site between the Palm and Fingers subdomains (Figure 2a; Cys423-Cys442 and Cys506-Cys509). The two disulfide bonds are found also in the crystal structures of Tpo, Tok and PMT DNA polymerases. Sequence alignment for archaeal DNA polymerases is shown in Figure 3, suggesting the potential for the formation of disulfide bonds in the same sites. It is thought that the disulfide bonds are required to maintain the structure of the Fingers and Palm subdomains at extremely high temperatures. Sequence comparison suggests that the number of disulfide bonds are correlated with optimum growth temperatures of organisms. DNA polymerases from *Thermococcus kodakarensis*, *Halobacterium volcanii* and *Archaeoglobus fulgidus*, with optimum growth temperatures of 88, 95 and 95°C, respectively, are expected to have one disulfide bond, because Cys469 is replaced by serine in *T. kodakarensis* and *H. volcanii*, and Cys467 is replaced by arginine in *A. fulgidus*. DNA polymerase from *Methanocaldococcus jannaschii*, with an optimum growth temperature of 65°C, is expected to have no disulfide bond, because Cys438, Cys442 and Cys536 are replaced by glutamic acid, arginine and serine, respectively.

Archaeal DNA polymerases have characteristic sequences of aromatic residues adjacent to glycine residues (Figure 3). These are localized at the hinges of the Palm subdomain at the exonuclease to the Fingers and Thumb-I subdomains (Figure 2a). These aromatic residues may provide a flexible aromatic environment because of the adjoining glycine residues. This may contribute

to the conformational changes of Pol domain in polymerization.

#### The 3'-5' exonuclease domain

DNA is synthesized by competition between the rate of polymerase and exonuclease activities at the newly synthesized 3' terminus from the primer. Misincorporation of a nucleotide destabilizes the structure of duplex DNA at the 3' terminus of the primer. This decreases the rate of nucleotide attack on the  $\alpha$ -phosphate group of the incoming dNTP by the primer 3'-OH and allows excision of the incorrect nucleotide by the proofreading exonuclease. The excision requires the movement of the 3' terminus to the exonuclease active site accompanied by unwinding of the duplex DNA, because the exonuclease active site is set apart from the polymerase active site. In KOD DNA polymerase, the exonuclease active site is set apart from the polymerase active site by approximately 40 Å. The editing complex of K869 DNA polymerase shows structural similarity to the editing mode of family B DNA polymerase.<sup>6</sup> The DNA polymerase binds the mismatched primer-template DNA, which is partially denatured, the 3' end of the primer strand is bound at the exonuclease site. Residues 251-262 of K869 DNA polymerase, that form an extended  $\beta$ -hairpin structure that sits directly on from the protein surface and projects into the DNA, stabilize the partially denatured or melted structure. Arg268 extending from the  $\beta$ -hairpin motif plays an important role. Arg268 and Phe123 appear to block the template strand by making interactions with the penultimate base at the 3' end of the primer-template. Arg268 and Phe123 in K869 DNA polymerase correspond to Arg267 and Phe151 in KOD DNA polymerase, respectively. Figure 2(b) shows the structural comparison of Exo domains of KOD and K869 DNA polymerases. Molecular surface and electrostatic potentials are shown in Figure 4. The  $\beta$ -hairpin motif in KOD DNA polymerase corresponds to residues 242-249 and Arg247, extending to the forked-point, which is the junction of the template-binding and editing clefts (T-cleft and E-cleft, respectively) (Figure 4). It seems that Arg247 can separate template strand from primer strand and stabilize the melted structure of the strands in a manner similar to that of the K869 DNA polymerase. As Phe152 is set apart from the active site, it is apparently unable to make an aromatic interaction with the base of the primer. Based on the above idea, the movement of the loop including Phe152 (Figure 2(b)) is required to interact with the primer strand of the E-cleft. Furthermore, Arg243 extends from the  $\beta$ -hairpin structure to the T-cleft. Arg243 interacts with the template strand to fix it at the T-cleft. In addition to Arg243 and Arg247, five arginine residues gather at the forked-point in KOD DNA polymerase (Arg245, Arg266, Arg346, Arg381 and Arg381) and provide a basic environment (Figures 2(b) and 4). It seems that they can interact with the phosphate



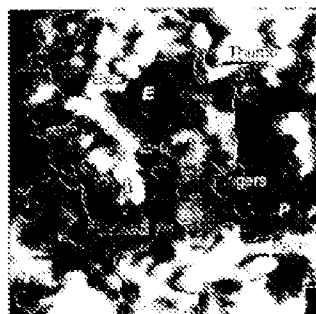


Figure 4. Molecular surface with electrostatic potential map around the helical pocket. The red and blue surfaces are acidic and basic regions, respectively. Donor and acceptor sites are labeled with orange letters. Hydroxyl and lysine/arginine active sites are labeled with 3' and 5', respectively. The  $\beta$ -strand is labeled with  $\beta$ .

#### Exon connection site

The KOD DNA polymerase gene encodes a 1671 amino acid residues precursor protein. The precursor protein is processed precisely into three parts by protein splicing. The self-splicing reaction yields the mature KOD polymerase (774 residues) and two intervening protein domains (termed introns), PI-Pol (360 residues) and PI-PolII (537 residues) as a result of the ligation of the external N and C-terminal domains (termed exons).<sup>10,15</sup> All known precursor proteins contain conserved serine acids at self-splicing sites: serine, threonine or cysteine (nucleophiles) at the intron N terminus, and His-Gln pair at the intron C terminus followed by serine, threonine, cysteine (nucleophilic) at the C-terminal N terminus.<sup>15</sup> The traces of the protein splicing reaction in KOD DNA polymerase are Ser497 and Ser498, which were knotted in at the N terminus of the C-exon. In the crystal structure of KOD DNA polymerase, the nucleophilic residues are located in the Pol domain (Figure 4a).

Nucleic acid sites in artificial family B DNA polymerases (in family) are classified into three types: Pol-Pol-1, Pol-Pol-1 and Pol-Pol-2 (The Protein Database, <http://www.ncbi.nlm.nih.gov/ncbi/interact.html>). The nucleophilic residues, serine or threonine, in the three sites are mapped in Figure 2(a). In the case of KOD DNA polymerase, PI-Pol intervenes in the Pol-Pol-1 site and PI-PolII intervenes in the Pol-Pol-2 site. The structure shows that they are localized around the polymerase active site in the Pol domain. Although they are exposed to solvent, they are surrounded by the

Fingers subdomain and the Thumb domain. The two introns cannot exist in the space because of steric hindrance. Therefore, it is necessary that the binding of introns and the subsequent self-excision are carried out before the exon is folded.

## Materials and Methods

### Crystallization

KOD DNA polymerase was overexpressed in *E. coli* BL21(DE3) and purified by the previously reported method.<sup>16</sup> The crystals of KOD DNA polymerase were grown by the previously reported method.<sup>17</sup> KOD DNA polymerase was concentrated up to about an  $A_{280}$  of 1.0. Crystals of KOD DNA polymerase suitable for diffraction experiments were obtained at 258 K with hanging drops of 2  $\mu$ l of protein solution and 2  $\mu$ l of reservoir solution, containing 160 mM sodium citrate buffer (pH 5.5) and 25 mM (v/v) 2-methyl-2,6-pyridinediol (MOPS), equilibrated against the reservoir solution.

### Data collection

X-ray diffraction measurements were performed at the beamline 18B of the Photon Factory at the High Energy Accelerator Research Organization, Tsukuba, Science City, Japan. Each crystal of KOD DNA polymerase was picked up directly with a nylon filter loop from a drop of mother liquid; the crystal was then rapidly transferred to the  $N_2$  gas stream. The incident beam with wavelength of 1.00 Å was attenuated to 0.2 mm in diameter. Intensity data were collected on 200 mm  $\times$  800 mm imaging plates (Fujifilm Company Ltd.) using the Marrowberg camera for macromolecules with a radius of 650 mm<sup>18,19</sup> and the collimator coated with 3° rotation per frame. The crystals diffracted at least to 2.5 Å resolution at 100 K. X-ray diffraction data were processed and sorted with programs DENZO and SCALEPACK.<sup>20</sup> The diffraction data were scaled with *scat* a cutoff threshold parameters were determined as  $a=111.9$  Å,  $b=111.4$  Å and  $c=73.9$  Å with the space group of  $P2_12_12_1$ . The unit-cell parameters gave Matthews's coefficient of 2.60 Å<sup>3</sup>/Da<sup>1</sup> and a solvent content of 52.2% (v/v).<sup>21</sup> The final completeness of the data consisted of 118,000 measurements of 36,798 unique observed reflections with an overall  $R_{\text{int}}$  of 8.4% and 34.3% in the outermost resolution shell (2.34–2.26 Å). This represents 95.1% of theoretically observable reflections at 2.3 Å resolution. The (outermost resolution shell) of data is 65.7% complete.

### Structure determination

The crystal structure of KOD DNA polymerase was solved by molecular replacement with the AMoRe program.<sup>22</sup> The structure of Tgo DNA polymerase (PDB code 1GGO) reduced to polynucleotide was used as the search model. Data in the resolution range of 3.0–2.5 Å were used in both the rotation and translation functions. Results are discussed in terms of the AMoRe correlation coefficient (CC). Using a Patterson solvent radius of 36 Å, a set of 20 rotation function peaks was observed, with the top peak having an AMoRe CC value of 13.4. The top rotation by translation function is CC of 45.1 with an  $R$ -factor of 54.1%. At this stage, the electron density of the Thumb domain is very ambiguous. Therefore, structural refinement of the initial stage was carried out with

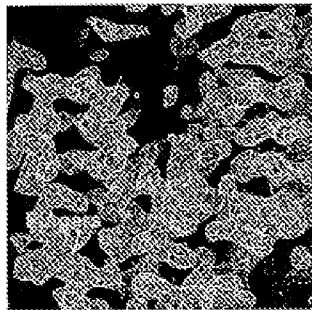


Figure 5. The final  $2F_o - F_c$  map around the Fingers and Palm subdomains. The map is contoured at 1  $\sigma$ .

a model containing the Thymine strand. The model was rationally modified using the program *Coot* and subjected to further rounds of refinement using data in the resolution range 40.6–3.0 Å with the program *REFMAC*.<sup>28</sup> The final  $R$ -factor is 23.1% and  $R_{free}$  is 31.3%, with  $R$ -factor deviations for bond lengths and bond angles being 0.007 Å and 5.1°, respectively. The 5f residues at the tip of each Thymine domain are not included in the final model due to poorly defined electron density. Figure 5 shows the final  $2F_o - F_c$  map superimposed on the refined 3D coordinates of KOD DNA polymerase.

#### Protein Data Bank accession code

Refined coordinates and structure factors have been deposited in the RCSB Protein Data Bank under the accession code 1OCX.

#### Figure preparation

Figures 1 and 2 were prepared using programs MOLSCRIPT<sup>29</sup> and Raster3D.<sup>30,31</sup> Figure 4 was prepared by GRASP.<sup>32</sup> Figure 5 was prepared using the program *Coot*.<sup>28</sup>

#### Acknowledgments

We thank Professor M. Sekiguchi, Dr. M. Watanabe, Dr. M. Suzuki and Dr. H. Igata for support in data collection at KEK-PF. This study was supported by TARA Sakurai Group of University of Tsukuba. The authors are grateful for a JSPS Fellowship for Japanese Junior Scientists.

#### References

1. Bradburne, G. K. & Rio, J. (1993). Comparative alignment and phylogenetic relationships of DNA polymerases. *Nucl. Acids Res.* 21, 767–802.
2. Klenow, R. E., Washington, W. T., Prasad, S. & Prasad, L. (1990). Strapping the gap: a family of novel DNA polymerases that replicate faulty DNA. *Proc. Natl. Acad. Sci. USA* 86, 12234–12239.
3. Klenow, R. E., Weaver, W. J. & Gorbach, V. L. (1993). The many faces of DNA polymerases: strategies for mutagenesis and for structural studies. *Proc. Natl. Acad. Sci. USA* 90, 3651–3653.
4. Wang, J., Satoh, A. K. M. A., Wang, C. C., Kerton, J. D., Klenow, R. E. & Sliemers, T. A. (1997). Crystal structure of a pol  $\alpha$  family replication DNA polymerase from *Haemophilus* HB99. *Cell* 89, 1087–1099.
5. Hagiwara, K.-P., Eickinger, A., Eick, S. A., Lasse, F., Aebihausen, W., Huber, R. & Angerer, R. (1999). Crystal structure of a thermophilic type B DNA polymerase from *Thermococcus* *sp.* *Proc. Natl. Acad. Sci. USA* 96, 2659–2665.
6. Skamene, Y. & Sliemers, T. A. (1998). Building a replication fork: interacting protein sliding along a template from DNA polymerase and a polynucleotide complex. *Cell* 94, 325–336.
7. Zhao, Y., Grollman, D., Adams, L., Leighton, L., Latham, R. & Kuriyan, J. (1998). Crystal structure of an archaeobacterial DNA polymerase. *Structure* 7, 1189–1199.
8. Rodriguez, A. C., Park, H.-W., Kato, C. & Sliemers, T. A. (2000). Crystal structure of a pol  $\alpha$  family DNA polymerase from the hyperthermophilic archaeon *Thermococcus* *sp.* *J. Mol. Biol.* 299, 471–482.
9. Fujimori, K., Ogasawara, S. & Imanaka, T. (1996). The world of archaea: genome analysis, evolution and thermophilic enzymes. *Gene* 179, 165–170.
10. Takagi, M., Nishioaka, M., Kikuchi, H., Nishioaka, M., Imanaka, T., Kawaiyama, H., Imai, M. & Imanaka, T. (1997). Characterization of DNA polymerase from *Pyrococcus* *sp.* strain KOD1 and its application to PCR. *Appl. Environ. Microbiol.* 63, 4804–4813.
11. Dujardin, S., Yano, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* 394, 251–256.
12. Li, Y., Kuroda, S. & Yoshimura, T. (1999). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus* *sp.* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* 17, 7534–7545.
13. Yang, G., Lin, Y.-C., Kato, S. & Klenow, R. E. (1999). Steady-state kinetic characteristics of KOD DNA polymerase mutants that affect dNTP incorporation. *Biochemistry* 38, 8366–8375.
14. Grayling, B. A., Sandman, K. & Roese, J. H. (1996). DNA stability and DNA binding proteins. *Annu. Rev. Biochem.* 65, 457–487.
15. Nishioaka, M., Fujimori, K., Takagi, M. & Imanaka, T. (1998). Characterization of two protein binding endonucleases associated in the DNA polymerase gene of *Pyrococcus* *sp.* *Nucl. Acids Res.* 26, 4409–4417.
16. Verier, P. B., Olson, G. J. & Adams, E. (1997). Comparison and analysis of protein sequences. *Nucl. Acids Res.* 25, 1087–1095.
17. Nishioaka, H., Matsumoto, T., Nishioaka, M., Yano, S., Taketake, H., Imai, T., Fujimori, K., Takagi, M.,

- Imanaka, T. & Kai, T. (1994). Crystallographic studies on a family B DNA polymerase from hyperthermophilic archaeon *Pyrococcus falciformis* strain KOD1. *J. Biol. Chem.* 269, 980-986.
28. Sakabe, M., Kozumi, S., Sukeike, K., Miyoshi, T., Nakagawa, A., Watanabe, N., Adachi, S. & Suzuki, S. (1998). Moselyberg camera for macromolecules with imaging plate data collection system at the photon factory: present status and future plan. *Rev. Sci. Instrum.* 69, 1226-1231.
29. Watanabe, N., Nakagawa, A., Adachi, S. & Sakabe, M. (1995). Macromolecular crystallography station BL-18B at the photon factory. *Rev. Sci. Instrum.* 66, 1824-1826.
30. Chwastowski, Z. & Minor, W. (1992). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 205, 307-326.
31. Ruppel, R. W. (1960). Sediment content of protein crystals. *J. Biol. Chem.* 23, 491-497.
32. Moras, J. (1994). *Adonis*: an automated package for molecular replacement. *Acta Crystallogr. sect. A*, 50, 157-162.
33. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in the models. *Acta Crystallogr. sect. A*, 47, 110-119.
34. Belanger, A. L., Adams, P. D., Chase, G. M., DiLiano, W. L., Goss, P., Grasse-Forstern, S. W., Jiang, J.-S., Kuznetsov, J., Niggli, M., Parnis, N. S., Read, R. J., Rice, L. M., Simonsen, T. & Warren, C. L. (1999). Crystallography & ISOLDE system: a new software suite for macromolecular structural determination. *Acta Crystallogr. sect. D*, 54, 908-921.
35. Kuszon, P. J. (1991). MOLSCREPT: a program to produce both density and electron phase of protein structures. *J. Appl. Crystallogr.* 24, 548-550.
36. Merritt, E. A. & Murphy, M. E. (1994). *StarSeries* version 2.0: a program for photomontage molecular graphics. *Acta Crystallogr. sect. D*, 50, 869-873.
37. Merritt, E. A. & Section, D. J. (1997). *Render3D* photomontage molecular graphics. *Methods Enzymol.* 227, 395-399.
38. Nicholls, A., Sharp, K. A. & Honig, B. (1991). GRASP: Protein folding and association: insights from the interfacial and electrostatic properties of lysozymes. *Protein: Struct. Funct. Genet.* 33, 281-296.

Edited by R. Huber

(Received 27 August 2000; received in revised form 8 December 2000; accepted 8 December 2000)

# Crystal structure of an archaeobacterial DNA polymerase

Yanxiang Zhao<sup>1†</sup>, David Jeruzalmi<sup>1†</sup>, Ismail Moarefi<sup>1,2</sup>, Lora Leighton<sup>1,2</sup>, Roger Lasken<sup>3</sup> and John Kurtyan<sup>1,2\*</sup>

**Background:** Members of the Pol II family of DNA polymerases are responsible for chromosomal replication in eukaryotes, and carry out highly processive DNA replication when attached to ring-shaped processivity clamps. The structures of Pol II polymerases are distinct from those of members of the well-studied Pol I family of DNA polymerases. The DNA polymerase from the extremophile *Thermoplasma acidophilum* strain TAP 20, Pol Pta is a member of the Pol II family that exhibits robust activity at elevated temperatures.

**Results:** The crystal structure of D. TAP Pol Pta has been determined at 2.8 Å resolution. The architecture of this Pol II-type DNA polymerase resembles that of the DNA polymerase from the bacteriophage RB69, with which it shares less than 20% sequence identity. As in RB69, the central catalytic region of the DNA polymerase is located within the "palm" subdomain and is structurally similar in structure to the corresponding regions of Pol I-type DNA polymerases. The structural scaffold that supports the catalytic core in D. TAP Pta is consistent in sequence to that of Pol I-type polymerases. The 3'-5' proofreading subdomain consists of D. TAP Pta resembles the corresponding domains of RB69 Pol and Pol I-type DNA polymerases. The exonuclease domain in D. TAP Pta is located on the opposite side of the palm subdomain compared to its location in Pol I-type polymerases. The 3' terminal domain of D. TAP Pta has structural similarity to DNA-binding domains. Sequence alignments suggest that this domain is conserved in the eukaryotic DNA polymerases  $\delta$  and  $\epsilon$ .

**Conclusions:** The solution of D. TAP Pol Pta confirms that the modes of binding of the template and synthesis of newly synthesized duplex DNA are likely to be similar in both Pol II and Pol I-type DNA polymerases. However, the mechanism by which the newly synthesized product transits in and out of the proofreading exonuclease domain has to be quite different. The discovery of a domain that seems to act as an RNA-binding module raises the possibility that Pol II family members interact with RNA.

**Address:** <sup>1</sup>Laboratories of Molecular Biophysics, Frederick Hughes Maclellan Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA and <sup>2</sup>Molecular Biology, Inc., 88 High Street, Canton, CT 06034, USA

<sup>3</sup>The two authors contributed equally to this work.

\*Corresponding author  
E-mail: kurtyan@maclellan.rockefeller.edu

**Key words:** *Thermoplasma acidophilum* TAP, DNA polymerase, Pol II family, DNA-binding domain, exonuclease, 3'-5' proofreading

Received: 12 April 2009

Revisions received: 27 May 2009

Revisions received: 2 June 2009

Accepted: 2 June 2009

Published: 29 September 2009

Domestic license: 1399, 7:1189-1195

cop: license: not available for RB69 Pol I and Pol II

ORCID: 212070618 - see your paper

© 2009 Kurtyan et al.; licensee Science Ltd. All rights reserved.

## Introduction

DNA polymerases can be classified into at least three families on the basis of sequence similarities to the three distinct DNA polymerases of *Escherichia coli*, Pol I, Pol II and Pol III [1]. Members of the Pol I family have been studied extensively, resulting in a comprehensive understanding of their functional properties and their structure [2-6]. In contrast to the detailed knowledge that is now available for the Pol I family, the Pol II and Pol III polymerases are poorly understood. The first crystal structure determined for a Pol II family member was that of the DNA polymerase of the bacteriophage RB69 (RB69 Pol) [7] and no structural information is currently available for any member of the Pol III family. Members of the Pol II (also known as Pol  $\delta$  or Pol  $\epsilon$ ) and Pol III families carry out processive replication of chromosomal DNA during cell division [8], and there is interest in further extending our

knowledge of their structures and mechanisms. Archaeobacterial DNA polymerases and the eukaryotic DNA polymerases  $\alpha$ ,  $\delta$  and  $\epsilon$  are members of the Pol II family [1].

The structure of RB69 Pol revealed that the general architecture of the core of the Pol II polymerases is strikingly similar to that of the Pol I polymerases [7]. Pol I polymerases are constructed from three smaller subdomains, termed the thumb, palm and fingers region by analogy to elements first noted in the structure of the Klenow fragment of *E. coli* DNA polymerase I [9]. In addition, Pol I DNA polymerases have a proofreading 3'-5' exonuclease domain located below the thumb subdomain, near the region where duplex DNA exits the polymerase active site [8,5]. Besides the residues involved in catalysis, there is no significant sequence similarity between the polymerase domains of members of the Pol I and Pol II families [1].



However, the subdomain architecture of the Pol I family is conserved in the RB69 structure, even though the detailed structures of the subdomains are quite divergent [7]. The structure domains of Pol I and Pol II DNA polymerases are closely related in sequence and, not surprisingly, the structure of the exonuclease domain of RB69 resembles that of the Pol I type polymerases. Given the general similarity in the polymerase domains of the Pol I polymerases and RB69, the location of the exonuclease domain in RB69 was a surprise. In RB69 the 3'-5' exonuclease domain is located above the fingers and opposite the thumb subdomains, suggesting that the shuffling of DNA between the polymerization and proofreading sites must occur by a different mechanism in Pol II DNA polymerase [7].

The mechanism of the Pol I family DNA polymerases is one unpaired in front [4,5,11,28]. The chemistry of nucleotide addition is mediated by two metal ions that are ligated by two separate residues. These are located in the palm subdomain, at the base of a deep cleft in the polymerase domain. High-resolution crystal structures of the Pol I type DNA polymerases of *T. thermophilus* (T<sub>7</sub> Pol) and *Thermus aquaticus* (Taq Pol) complexed to pyrimidine-nucleoside (PNA) and incoming nucleotide have been determined, allowing the mechanism of nucleotide incorporation and selectivity to be visualized [10,11,28]. Although corresponding structural information for the Pol II family DNA polymerases is lacking, antibodies in general organization of the polymerase core as well as sequence conservation within critical elements of the central palm subdomain suggest that general features of the recognition of DNA will be similar in Pol II polymerases.

The DNA polymerase from the archaebacterium *Sulfolobus* strain T<sub>7</sub> (D. T<sub>7</sub> Pol) is a member of the Pol II family, and has both the exonuclease DNA polymerase and 3'-5' exonuclease activities [12]. D. T<sub>7</sub> Pol contains undiminished DNA polymerase activity after incubation at 90°C for one hour (RL, unpublished results). The sequence of D. T<sub>7</sub> Pol is very closely related (ca 70% identity) to that of other archaebacterial DNA polymerases, such as those from *Pyrococcus furiosus* [13] and *Thermococcus kodakarensis* [14]. D. T<sub>7</sub> Pol is also related to eukaryotic DNA polymerases  $\alpha$ ,  $\delta$  and  $\epsilon$  (54% sequence identity over 95% residues of the DNA polymerase core for the human  $\delta$  sequence [1]). The archaebacterial genomes also contain genes coding for proteins with close homology to proliferating cell nuclear antigen (PCNA), the DNA polymerase clamp in eukaryotes, as well as subunits of the clamp-loader molecule PCF-C (replication factor C). It is likely that archaebacterial DNA polymerases achieve processivity by attachment to the ring-shaped PCNA ring, although direct evidence for such a mechanism is lacking.

We have determined the structure of D. T<sub>7</sub> Pol at 2.6 Å resolution. D. T<sub>7</sub> Pol shares less than 30% sequence

identity with RB69 Pol, but the structures of the two enzymes resemble each other closely. The structure reported here has been determined to the precision of DNA. Nevertheless, the close structural correspondence between the active sites of Pol I and Pol II DNA polymerases allows inferences to be made about the mode of DNA recognition by D. T<sub>7</sub> Pol. The very N-terminal region of D. T<sub>7</sub> Pol encodes a domain (residues 1-123) that is closely related to conserved single-stranded RNA-binding domains (RRDs), also known as RNA-recognition motifs (RRMs) [15]. The structure of the 3'-5' proofreading exonuclease domain of D. T<sub>7</sub> Pol is similar to those of the Pol I type polymerases. However, its location relative to the palm subdomain resembles the location seen in RB69 [7] rather than the Pol I type polymerases [9,16,17]. The structure of D. T<sub>7</sub> Pol reported here provides further evidence that the mode of DNA-template recognition and the distinct editing channel established for the Pol II family by the presence of RB69 Pol is valid for the entire Pol II family.

## Results and discussion

### Structure determination

Crystals of D. T<sub>7</sub> Pol have been obtained from 2,4-methylenediphenol (MPPD) (Maeve II) and polyethylene glycol (PEG) (Native II). Both crystal forms are orthorhombic (P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>,  $a = 64.6$  Å,  $b = 107.6$  Å,  $c = 151.1$  Å for Native I and  $a = 68.1$  Å,  $b = 89.6$  Å,  $c = 155.9$  Å for Maeve II). Experimental phases (Table 1) at 3.0 Å were obtained from four isomorphous heavy-atom derivatives, using Native II and the program SHARP [17]. Phases were improved by iterative cycles of real space density modification, consisting of solvent flattening and negative density truncation, using SOLIDMOD [18,19]. The resulting electron-density map allowed the chain to be traced unambiguously, with ready determination of sequence register. The model was refined to 2.6 Å against data for Native II (R value = 24.2%,  $R_{\text{free}}$  = 29.5%) and subsequently to 2.3 Å against data for Native I (R value = 25.3%,  $R_{\text{free}}$  = 29.9%), using CNS [20]. The model for Native II is somewhat more complete (see the Materials and methods section) and is used for most of the discussion. This model includes 746 residues from 1 to 736 in Native II. Amino acids 386-389 and 605-616 are not visible in our electron-density maps and are not included in the model.

### General description of the structure

D. T<sub>7</sub> Pol (Figure 1) is composed of a polymerase domain (residues 260-773) and an exonuclease domain (residues 133-385), as well as an N-terminal domain (residues 1-123) that is not found in Pol I type DNA polymerases [4]. The polymerase domain is further comprised of three smaller subdomains, termed the thumb (residues 667-758), palm (residues 390-443) and fingers (residues 446-499). The structures of the MPPD and PEG400 crystal forms of D. T<sub>7</sub> Pol are very similar

Table 3

Data collection, structure determination and refinement statistics.

	Resolution (Å)	Number of reflections (unique)	Completeness (%)	$R_{\text{int}}^a$ (%)	$R_{\text{pact}}^b$ (%)	Scale	Phasing power <sup>c</sup>	Figure of merit <sup>d</sup>
Native data	—	—	—	—	—	—	—	—
Native 1	30.0–2.8	32,000	92.0(17.8)	8.2(5.6)	—	—	—	—
Native 2	30.0–2.8	42,840	97.2(23.4)	4.9(3.1)	57.4	—	—	—
MTLS analysis	—	—	—	—	—	—	—	0.267
$\sigma_1$	50.0–3.0	40,310	97.9(23.0)	8.4(2.0)	10.1	4	1.34(0.00)	0.274
$\sigma_2$	50.0–3.0	34,008	96.2(19.4)	8.8(8.8)	13.6	1	1.18(0.00)	0.185
$\sigma_3$	50.0–3.0	55,107	89.0(27.4)	9.6(2.5)	10.1	5	1.30(0.00)	0.327
Refinement	—	Number of reflections ( $R_f = 2\sigma$ )	—	$R_{\text{pact}}^b$ ( $R_{\text{pact}}^b$ ) (%)	Total number of atoms	Time by hours (d)	Time by minutes (s)	Time by minutes (s)
Native 1	30.0–2.8	31,591	24.3(29.1)	3.1(1)	107,982(7)	1:01:58(1)	1:01:58(1)	1:01:58(1)
Native 2	30.0–2.8	37,173	17.5(20.8)	4.3(1)	108,022(7)	1:02:10(1)	1:02:10(1)	1:02:10(1)

$R_{\text{pact}}^b = 100 \times \sum |F_o - F_c| / \sum F_o$ , where  $F_o$  is the observed intensity of a given reflection,  $F_c$  is the calculated intensity of a given reflection,  $F_o$  and  $F_c$  are the observed and calculated structure factor amplitudes, respectively.

<sup>c</sup>Phasing power =  $\sum |F_o - F_c| / \sum |F_o|$ , where  $F_o$  is the observed intensity of a given reflection,  $F_c$  is the calculated intensity of a given reflection.

<sup>d</sup>Figure of merit =  $\sum |F_o - F_c| / \sum |F_o|$ , where  $F_o$  is the observed intensity of a given reflection,  $F_c$  is the calculated intensity of a given reflection.

<sup>e</sup>Phasing power =  $\sum |F_o - F_c| / \sum |F_o|$ , where  $F_o$  is the observed intensity of a given reflection,  $F_c$  is the calculated intensity of a given reflection.

<sup>f</sup>Phasing power =  $\sum |F_o - F_c| / \sum |F_o|$ , where  $F_o$  is the observed intensity of a given reflection,  $F_c$  is the calculated intensity of a given reflection.

<sup>g</sup>Phasing power =  $\sum |F_o - F_c| / \sum |F_o|$ , where  $F_o$  is the observed intensity of a given reflection,  $F_c$  is the calculated intensity of a given reflection.

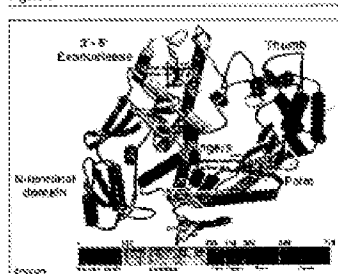
in terms of the individual subunits. The major difference between the two structures is a rotation of  $\sim 60^\circ$  in the orientation of the exonuclease domain with respect to the thumb subdomain.

The domains of  $\Omega$  Tdk Pol are arranged as an irregularly shaped flattened ring with a central cavity located near the polymerase active site. The mostly  $\alpha$ -helical thumb subdomain forms one side of the active-site cleft and makes contacts with the exonuclease domain (Figure 1). The structure of the thumb domains of various polymerases are often conserved in structure. However, in all cases where structures are available the thumb domain is seen to hold an important role by binding substrates with duplex DNA as it nears the polymerase active site [24]. The  $\Omega$  Tdk Pol structure has been determined in the absence of DNA, and a portion of the thumb subdomain that is likely to contact DNA (residues 665–675) is disordered. This is commonly observed for the corresponding regions of other polymerases in the absence of substrates [24–26]. In the DNA polymerase from *Thermoplasma* [24 and [26], the thumb subdomain also provides a conserved element that interacts with the processivity clamp [27,28]. In  $\Omega$  Tdk Pol, the corresponding region (residues 757–774) is disordered.

The central region of the active-site cleft is occupied by the palm subdomain and includes residues important for substrate discrimination and the catalysis of the polymerase reaction. In  $\Omega$  Tdk Pol, the palm is organized around three  $\beta$  strands ( $\beta$ 1b,  $\beta$ 1c,  $\beta$ 2b) flanked by an  $\alpha$  helix ( $\alpha$ Q) (Figures 1,2a,3a). It contains two disulfide

bonds (Cys438–Cys442, Cys436–Cys450) that have not been previously observed in palm subdomains and which may be important for thermostability (Figure 1).

Figure 1



Structure of  $\Omega$  Tdk Pol. The structure is represented by spheres for atoms, arrows for helices, and a flat ribbon for other secondary structure elements. Two grey spheres represent metal ions (potassium or sodium) bound to the exonuclease domain. The active site of the polymerase is marked by the presence of two water molecules (H<sub>2</sub>O and H<sub>2</sub>O<sub>2</sub>). The two disulfide bonds are indicated. Regions of the polypeptide chain that could not be modeled in the given conformational domain of disorder are indicated by dotted lines. The various domains and subdomains and their boundaries are indicated in the text.



of Pol I polymerases that are essential for isoenzymatic activity because they coordinate two metal ions [2,10,11,27]. The corresponding residues in D. Tok Pol are Asp688 and Asp643 (Figure 1). No metal ions are, however, visible in our electron-density maps.

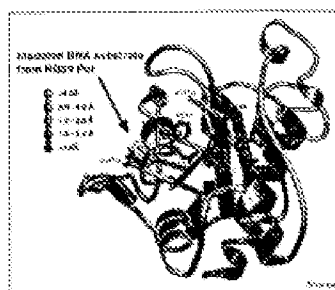
The fingers subdomain in D. Tok Pol contains 16's set of antiparallel  $\alpha$  helices ( $\alpha 10$ ,  $\alpha 11$ ,  $\alpha 12$ , Figure 2). These helices are shorter in length than the corresponding elements of RB69 Pol, and a helical segment that connects helices Q and N in RB69 Pol is missing altogether (Figure 3). The fingers domain of D. Tok Pol is unrelated in overall structure to that of Pol I type polymerases (Figure 2). However, helix  $\alpha 10$  in D. Tok Pol is positioned similarly to helix Q in Pol I polymerases (Figure 3), and is likely to play an analogous and crucial role in recognition of the incoming nucleotide [8-11,28].

The 3'-5' exonuclease domain in D. Tok Pol is located opposite the thumb subdomain and above the fingers subdomain, as noted for RB69 Pol. It contains two metal ions (presumably  $Mg^{2+}$ ) ligated to Asp141 and Glu143 (Figure 1). The position of this domain relative to the polymerase active site is distinct from the arrangement seen in Pol I type polymerases. The conservation between RB69 and D. Tok Pol at the location of the exonuclease domain suggests that this is a characteristic feature of Pol II type polymerases. The structure of the D. Tok Pol 3'-5' exonuclease domain resembles those associated with other DNA polymerases [29,30]. The 3'-5' exonuclease domains from the Pol I (*E. coli*, *T. aquaticus*, *Bacteriophage lambda* bacteriophage T7) or Pol II (RB69) polymerase families can be aligned over each other closely (data not shown) for around [118, 111, 112, 114 and helices  $\alpha 8$  and  $\alpha 1$  is in the range of 1.0-2.8 Å. This alignment superimposes residues associated with substrate binding, catalysis and metal binding in a satisfactory manner (Figure 4) [4].

The arrangement of the N-terminal, exonuclease, and polymerase domains creates two deep grooves leading into and out of the polymerase active site. The D groove (for duplex-DNA binding, following the noncatalytic site of [7]) is located immediately below the thumb subdomain and includes a region of positive electrostatic potential. The T groove (for template-DNA binding) leads away from the active site in the opposite direction and is located below the fingers subdomain. A small channel (the ridding channel) leads from the polymerase domain to the exonuclease active site (Figure 2).

We have used the structure of T7 Pol bound to primer-template DNA to model DNA into D. Tok Pol (Figure 2). Superposition of the palm subdomains of the two polymerases shows that remarkably few base contacts are formed between the DNA (from T7 Pol) and amino in the D. Tok Pol model. The  $\alpha$  region that does collide

Figure 3



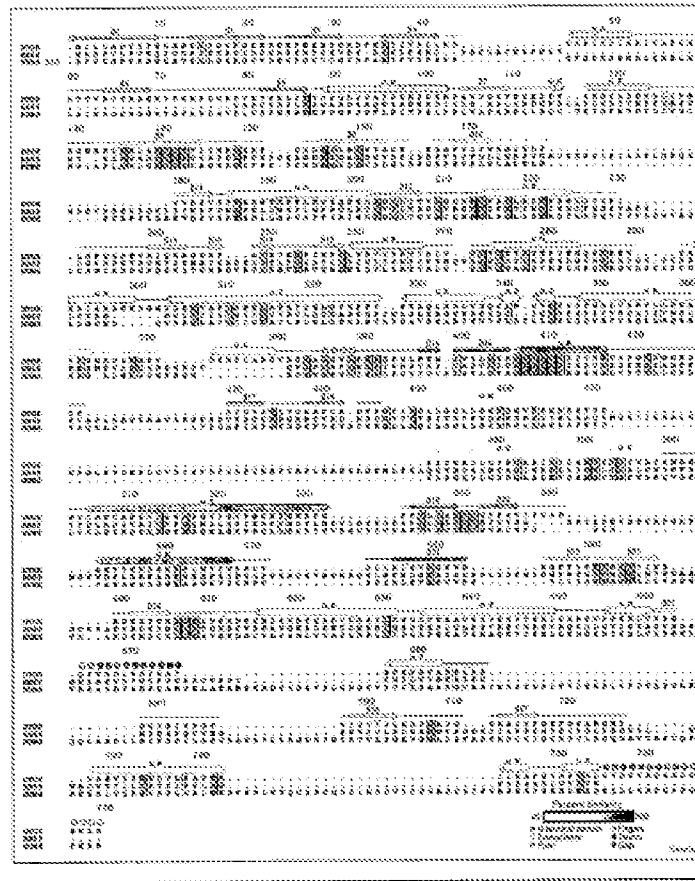
Structural alignment of exonuclease domains. Sequence of exonuclease domains from RB. RB69, 1775-1869, and D. Tok, 1521-1612, have been superimposed by superimposing residues 137-143, 148-154, 161-167, 200-206, 257-263, and 301-307, each equivalent region [118, 111, 112, 114] over each other. A color gradient is used to display the average error for the family of superimposed structures ranging from blue (1.0-1.5 Å) to white (2.0-2.5 Å). Residues conserved amongst prokaryotic sequences and implicated in catalysis are shown in green ball and stick representation. Two grey spheres represent two metal ions located in the active site. The active site is also indicated by a nucleotide (in grey) derived from superposition of the exonuclease domain from the RB69 Pol structure.

with the DNA is the segment connecting the exonuclease and polymerase domains. This region (residues 372-381) is partially disordered in the D. Tok Pol structure, and is likely to reorganize upon binding DNA. This superposition allows five base pairs of DNA to be accommodated in the D. Tok Pol active site, with the formation of DNA-protein contacts. The formation of contacts with additional base pairs would require a change in the position of the thumb subdomain in the region of the D groove. A change in the conformation of the fingers subdomain (helices  $\alpha 6$  and  $\alpha 7$ ) is also required to position residues Lys487 and Thr491 (or Tyr491) of D. Tok Pol (Figure 2) for interaction with the incoming nucleotide, by analogy with the T7 Pol structure [11]. Finally, the superimposed primer-template DNA is well positioned so that the incoming template strand will probably reside in the T groove. Superposition of the DNA molecule derived from the structure of RBV-1 reverse transcriptase superimposed to DNA [11] leads to similar conclusions.

#### Comparison between D. Tok Pol and RB69 Pol

Although the DNA polymerases from D. Tok Pol and bacteriophage RB69 share less than 30% primary sequence identity (Figure 2), their structures resemble each other

Figure 3



**Figure 5**

Sequence-based sequence alignment for D. Tok Pol (210aa), R869 Pol (200aa) and human Pol  $\beta$  (344aa). The R869 Pol sequence region at residue 170, as indicated by the number in the beginning of the sequence. The alignment is colored by sequence identity: 100%, white; 75-99%, green; calculated as described in Figure 1c. Shown here is a small subset of a larger set of sequences that were used to generate the alignment. The full sequence alignment is available in Supplementary Appendix 1. The respective secondary structural elements located at in Figure 1 are represented by helices as cylinders, strands as sheets, and loops as thin lines. Entry traces represent pointers to the corresponding chain that contain the full dataset.

clearly (Figure 1). Not surprisingly, the regions of highest sequence similarity are concentrated in and around the exonuclease and polymerase active sites (Figures 2a,b). Despite the low overall sequence identity, the individual subdomains in the two structures superimposed well (true root in C $\alpha$  position in the fingers, thumb and palm subdomains is in the range of 0.8 to 3.5 Å). Moreover, the overall arrangement of domains and subdomains with respect to each other is preserved in the two polymerases, underscoring the structural similarity of Pol  $\beta$  DNA polymerases share a common architecture (Figure 3).

One difference between the overall structure of D. Tok Pol and R869 Pol concerns the orientation of the exonuclease domain with respect to the rest of the structure. When the two polymerases are superimposed on their respective palm subdomains it is seen that the exonuclease domain of R869 is rotated towards by  $\sim 8^\circ$ , leaving the active site in a solvent-inaccessible configuration [2]. In contrast, the exonuclease domain in D. Tok Pol has its active site conventionally exposed to solvent. It is possible that conformational changes between open and closed configurations of the exonuclease domain are a part of the functional cycle

of the protein, particularly as the two different forms of D. Tok Pol differ in the orientation of the exonuclease domain (see above).

One interesting difference between D. Tok Pol and R869 Pol is that the former is a thermostable DNA polymerase whereas the latter is not. Unsurprisingly, attempts to identify features in the D. Tok Pol structure that might be correlated with thermostability is complicated by the very low sequence identity between the two sequences. One feature that does stand out, however, is the increased formation of arrays of ionic interactions on the surface of D. Tok Pol when compared to that of R869 Pol (Figure 6). The formation of networks of ionic interactions has been shown to correlate with thermostability in other proteins [16,32,33].

Generally, D. Tok Pol subdomains tend to be more compact with smaller helices and shorter loops than are found in R869 Pol, a feature that may be another important source of thermostability. For example, the palm subdomain displays close structural conservation of elements near the catalytic sequence residues. However, helix  $\alpha 2$  in D. Tok Pol is much shorter than its counterpart in R869 Pol, and a small subdomain in front of the palm subdomain is entirely missing in D. Tok Pol (Figures 3,8). Deletion of these elements is also seen in a representative set of archaeobacterial DNA polymerases [13,14]. Likewise, the finger subdomain is missing a large mass of loops in D. Tok Pol (Figure 3,8). However, the R869 finger subdomain most probably plays a T4 phage specific role, as it is also missing from one alignment of archaeobacterial DNA polymerases and eukaryotic polymerases [14] (Figure 5).

**The N-terminal domain resembles RNA-binding domains.**  
The N-terminal domain of D. Tok Pol has no corresponding element in Pol  $\beta$  type polymerases. Analysis of the

**Figure 6**

Comparison of surface charges in D. Tok Pol and R869 Pol. Accessible surface representation of full D. Tok Pol and full R869 Pol are side-by-side. The superposition of their palm subdomains. Surface regions corresponding to the exonuclease domain of exonuclease and polymerase are colored red, whereas surface regions corresponding to the catalytic region of fingers and thumb are colored blue. D. Tok Pol has a striking pairing of oppositely charged residues not seen in R869 pol. A representation of D. Tok Pol at a lower resolution for comparison.

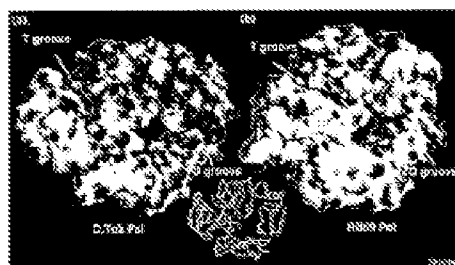
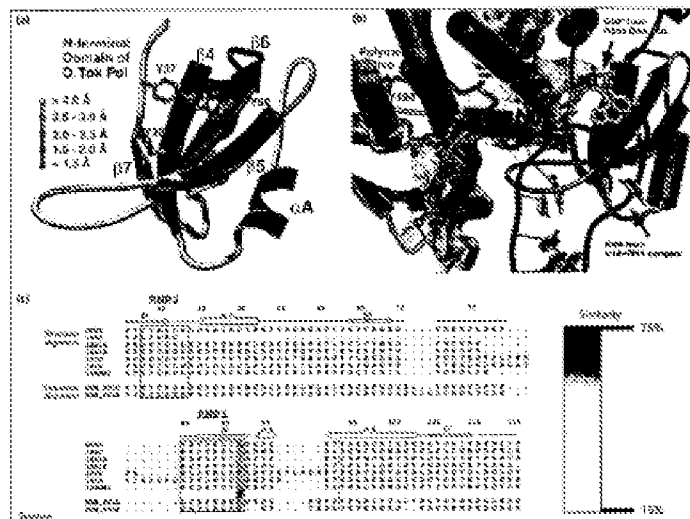


Figure 2



The 5'-noncoding region of the 30S ribosomal subunit contains a 16S ribosomal domain, a structure of 850 nucleotides from 3142 to 3991 nucleotides. It is flanked by 23S and 16S ribosomal genes. It has been shown to regulate 30S ribosome assembly [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181,182,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,208,209,210,211,212,213,214,215,216,217,218,219,220,221,222,223,224,225,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,243,244,245,246,247,248,249,250,251,252,253,254,255,256,257,258,259,260,261,262,263,264,265,266,267,268,269,270,271,272,273,274,275,276,277,278,279,280,281,282,283,284,285,286,287,288,289,290,291,292,293,294,295,296,297,298,299,300,301,302,303,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,330,331,332,333,334,335,336,337,338,339,340,341,342,343,344,345,346,347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,371,372,373,374,375,376,377,378,379,380,381,382,383,384,385,386,387,388,389,390,391,392,393,394,395,396,397,398,399,400,401,402,403,404,405,406,407,408,409,410,411,412,413,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428,429,430,431,432,433,434,435,436,437,438,439,440,441,442,443,444,445,446,447,448,449,450,451,452,453,454,455,456,457,458,459,460,461,462,463,464,465,466,467,468,469,470,471,472,473,474,475,476,477,478,479,480,481,482,483,484,485,486,487,488,489,490,491,492,493,494,495,496,497,498,499,500,501,502,503,504,505,506,507,508,509,510,511,512,513,514,515,516,517,518,519,520,521,522,523,524,525,526,527,528,529,530,531,532,533,534,535,536,537,538,539,540,541,542,543,544,545,546,547,548,549,550,551,552,553,554,555,556,557,558,559,560,561,562,563,564,565,566,567,568,569,570,571,572,573,574,575,576,577,578,579,580,581,582,583,584,585,586,587,588,589,590,591,592,593,594,595,596,597,598,599,600,601,602,603,604,605,606,607,608,609,610,611,612,613,614,615,616,617,618,619,620,621,622,623,624,625,626,627,628,629,630,631,632,633,634,635,636,637,638,639,640,641,642,643,644,645,646,647,648,649,650,651,652,653,654,655,656,657,658,659,660,661,662,663,664,665,666,667,668,669,670,671,672,673,674,675,676,677,678,679,680,681,682,683,684,685,686,687,688,689,690,691,692,693,694,695,696,697,698,699,700,701,702,703,704,705,706,707,708,709,710,711,712,713,714,715,716,717,718,719,720,721,722,723,724,725,726,727,728,729,730,731,732,733,734,735,736,737,738,739,740,741,742,743,744,745,746,747,748,749,750,751,752,753,754,755,756,757,758,759,760,761,762,763,764,765,766,767,768,769,770,771,772,773,774,775,776,777,778,779,780,781,782,783,784,785,786,787,788,789,790,791,792,793,794,795,796,797,798,799,800,801,802,803,804,805,806,807,808,809,810,811,812,813,814,815,816,817,818,819,820,821,822,823,824,825,826,827,828,829,830,831,832,833,834,835,836,837,838,839,840,841,842,843,844,845,846,847,848,849,850,851,852,853,854,855,856,857,858,859,860,861,862,863,864,865,866,867,868,869,870,871,872,873,874,875,876,877,878,879,880,881,882,883,884,885,886,887,888,889,890,891,892,893,894,895,896,897,898,899,900,901,902,903,904,905,906,907,908,909,910,911,912,913,914,915,916,917,918,919,920,921,922,923,924,925,926,927,928,929,930,931,932,933,934,935,936,937,938,939,940,941,942,943,944,945,946,947,948,949,950,951,952,953,954,955,956,957,958,959,960,961,962,963,964,965,966,967,968,969,970,971,972,973,974,975,976,977,978,979,980,981,982,983,984,985,986,987,988,989,990,991,992,993,994,995,996,997,998,999,1000,1001,1002,1003,1004,1005,1006,1007,1008,1009,1010,1011,1012,1013,1014,1015,1016,1017,1018,1019,1020,1021,1022,1023,10

[illegible]

analysis of this dataset using DAVID (38) (<http://www.ebi.ac.uk/daavid/>) revealed a previously unappreciated similarity to RBDs. RBDs are small modular

80-90 residues found in RNA-binding proteins of prokaryotes, eukarya, and eukaryotes (reviewed in [25]). These modules adopt a conserved  $\beta\alpha\beta\alpha\beta$  architecture

and bind to single-stranded RNA. Two conserved sequence motifs, referred to as RNP1 (phosphotransferase 1) and RNP2, provide acidic and charged residues that are important for RNA recognition [55] (Figure 7).

The 6-terminal domain of D. Tsk Poi can be superimposed closely onto the pore secondary structural elements of R80s by the C1A superpositioned domain (32). ribosomal protein 26 (33), the heterotrimeric ribulose-1,5-bisphosphate (RuBP) carboxylase from R80 (domains 153, 268) and the superoxide-binding domain from *E. coli* hemophyllin-RNA oxidase (324). The errors in C1A positions for these superpositions are in the range of 4.5–7.0 Å (Figure 7a). Correspondence between the coordinates of the larger in the 6-terminal domain of D. Tsk Poi and those of the RNA-binding domains are within the range of structural variation seen in the various RNA-binding domains.

There is no evidence at present to suggest that the N-terminal domain of D. T<sub>2</sub> P<sub>2</sub> binds Zn<sup>2+</sup>. However, comparisons with the sequence of the Zn<sup>2+</sup> complexes of the RNA-binding domain suggest that the N-terminal domain might in fact be a functional RNA-binding domain. The N-terminal domain of T<sub>2</sub> P<sub>2</sub> and T<sub>2</sub> P<sub>1</sub> are identical in the first 100 amino acids, whereas the amino acid sequence of the N-terminal domain of T<sub>2</sub> P<sub>2</sub> and T<sub>2</sub> P<sub>1</sub> differ in the last 100 amino acids, with the N-terminal domain of T<sub>2</sub> P<sub>2</sub> being more similar to the N-terminal domain of the RNA-binding domain of the T<sub>2</sub> P<sub>2</sub> than to the N-terminal domain of the RNA-binding domain of the T<sub>2</sub> P<sub>1</sub>. The N-terminal domain of T<sub>2</sub> P<sub>2</sub> and T<sub>2</sub> P<sub>1</sub> are identical in the first 100 amino acids, whereas the amino acid sequence of the N-terminal domain of T<sub>2</sub> P<sub>2</sub> and T<sub>2</sub> P<sub>1</sub> differ in the last 100 amino acids, with the N-terminal domain of T<sub>2</sub> P<sub>2</sub> being more similar to the N-terminal domain of the RNA-binding domain of the T<sub>2</sub> P<sub>2</sub> than to the N-terminal domain of the RNA-binding domain of the T<sub>2</sub> P<sub>1</sub>.

There is no significant overall sequence conservation between the N-terminal domain of 1, 2, 3, 4, 6, 8 and 9 RNA-binding domains, which is why the presence of this fold are not taxonomic parameters (Figure 7a). Comparison of the structures of 1 and 2 shows that the DNA phosphate and human polymerase  $\beta$  and a negatively charged  $\alpha$  corresponding structural element is likely to be found in these polymerases as well (Figure 7c). The sequence alignments in this region is unambiguous for the archaeobacterial DNA polymerases. For eukaryotic polymerases the alignments is less certain, but it is clear to conceive the essential catalytic character of the R258' motifs (Figure 7c). Conformation of the presence of these domains along with their ability to bind RNA, and their precise role in eukaryotic DNA synthesis awaits future structural and functional studies.

### ପିଲାପିଲାଙ୍କର ନ୍ୟାୟିକତା

The structure of the DNA polymerase from the archaeobacterium *Desulfurococcus* strain Tok. D. Tok Pol reveals a strong similarity in the DNA polymerase from bacteriophage T4693. It also reveals the presence of an  $\alpha$ -terminal domain that has structural similarity to DNA-binding domains from the T4A replication protein, ribosomal protein S6, the HIV-1 p66 protein and the archaeal-binding domain from 7' thermophilic phage- $\lambda$ CRP3. Although the structure in the immediate vicinity of the central catalytic region of the polymerase domain closely resembles that of Pol I type DNA polymerases, the overall architecture of Tok. D. Tok Pol and the placement of the exonuclease domain is strikingly different. The similarity between Tok. D. Tok Pol and HIV-1 Pol suggests that these two structures are representative of a new class of DNA polymerase fold. Members of this family carry out chromosomal DNA replication in eukaryotes, including humans, and yet there is no structural information available for any eukaryotic member of this family. While this monomerology was being prepared, the structure of another archaeobacterial DNA polymerase, that from the organism *Thermococcus garnettii*, has been reported [36]. The Tok. D. Tok Pol structure reported here, along with the HIV-1 Pol structure and the structure of the *Thermococcus garnettii* DNA polymerase, should now make it possible to generate reliable structural models for eukaryotic DNA polymerases.

## Materials and methods

[illegible][illegible]



0080) above  $2\theta = 55.1^\circ$ ,  $2\theta = 107.0^\circ$ ,  $2\theta < 135.0^\circ$ ,  $2\theta > 180^\circ$ ,  $2\theta < 180^\circ$ . Phase-contrast images were obtained by summing several images in stacking series containing 10 diffraction images reconstructed for  $2\theta$ .

#### Data collection and phase determination

X-ray diffraction data were from a set of shock-compressed metal and intermetallic alloy-alloy specimens were recorded at the Cornell High Energy Synchrotron Source (CHESS) position 61.3,  $\lambda = 0.0833$ . Data from frame 1 (frames 1 through 100) recorded to a single crystal of  $2\theta$ ,  $2\theta < 180^\circ$ ,  $2\theta < 180^\circ$  images was recorded at  $2\theta$ ,  $2\theta < 180^\circ$  (frames 101 through 1000), which yielded data to beyond  $2\theta = 180^\circ$ . X-ray diffraction data were recorded, integrated and scaled using the *XD4* package [43].

The process of data collection was followed manually by inspection of electron images and checked by computerized difference Fourier maps. Experimental phase were calculated using these data with the program *SHARP* [44]. In our hands, single crystal electron diffraction maps were obtained by performing individual single electron images, reconstructing (SIR) calculations in *SHARP* and combining the individual SIR images with using the program *SHARP* [45]. The experimental data were integrated and checked by solvent mapping and negative density projection as implemented in *SHARP*. The procedure *SHARP* [46] [47] [48] [49] [50] [51] [52] [53] [54] [55] [56] [57] [58] [59] [60] [61] [62] [63] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [80] [81] [82] [83] [84] [85] [86] [87] [88] [89] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100] [101] [102] [103] [104] [105] [106] [107] [108] [109] [110] [111] [112] [113] [114] [115] [116] [117] [118] [119] [120] [121] [122] [123] [124] [125] [126] [127] [128] [129] [130] [131] [132] [133] [134] [135] [136] [137] [138] [139] [140] [141] [142] [143] [144] [145] [146] [147] [148] [149] [150] [151] [152] [153] [154] [155] [156] [157] [158] [159] [160] [161] [162] [163] [164] [165] [166] [167] [168] [169] [170] [171] [172] [173] [174] [175] [176] [177] [178] [179] [180] [181] [182] [183] [184] [185] [186] [187] [188] [189] [190] [191] [192] [193] [194] [195] [196] [197] [198] [199] [200] [201] [202] [203] [204] [205] [206] [207] [208] [209] [210] [211] [212] [213] [214] [215] [216] [217] [218] [219] [220] [221] [222] [223] [224] [225] [226] [227] [228] [229] [230] [231] [232] [233] [234] [235] [236] [237] [238] [239] [240] [241] [242] [243] [244] [245] [246] [247] [248] [249] [250] [251] [252] [253] [254] [255] [256] [257] [258] [259] [260] [261] [262] [263] [264] [265] [266] [267] [268] [269] [270] [271] [272] [273] [274] [275] [276] [277] [278] [279] [280] [281] [282] [283] [284] [285] [286] [287] [288] [289] [290] [291] [292] [293] [294] [295] [296] [297] [298] [299] [300] [301] [302] [303] [304] [305] [306] [307] [308] [309] [310] [311] [312] [313] [314] [315] [316] [317] [318] [319] [320] [321] [322] [323] [324] [325] [326] [327] [328] [329] [330] [331] [332] [333] [334] [335] [336] [337] [338] [339] [340] [341] [342] [343] [344] [345] [346] [347] [348] [349] [350] [351] [352] [353] [354] [355] [356] [357] [358] [359] [360] [361] [362] [363] [364] [365] [366] [367] [368] [369] [370] [371] [372] [373] [374] [375] [376] [377] [378] [379] [380] [381] [382] [383] [384] [385] [386] [387] [388] [389] [390] [391] [392] [393] [394] [395] [396] [397] [398] [399] [400] [401] [402] [403] [404] [405] [406] [407] [408] [409] [410] [411] [412] [413] [414] [415] [416] [417] [418] [419] [420] [421] [422] [423] [424] [425] [426] [427] [428] [429] [430] [431] [432] [433] [434] [435] [436] [437] [438] [439] [440] [441] [442] [443] [444] [445] [446] [447] [448] [449] [450] [451] [452] [453] [454] [455] [456] [457] [458] [459] [460] [461] [462] [463] [464] [465] [466] [467] [468] [469] [470] [471] [472] [473] [474] [475] [476] [477] [478] [479] [480] [481] [482] [483] [484] [485] [486] [487] [488] [489] [490] [491] [492] [493] [494] [495] [496] [497] [498] [499] [500] [501] [502] [503] [504] [505] [506] [507] [508] [509] [510] [511] [512] [513] [514] [515] [516] [517] [518] [519] [520] [521] [522] [523] [524] [525] [526] [527] [528] [529] [530] [531] [532] [533] [534] [535] [536] [537] [538] [539] [540] [541] [542] [543] [544] [545] [546] [547] [548] [549] [550] [551] [552] [553] [554] [555] [556] [557] [558] [559] [560] [561] [562] [563] [564] [565] [566] [567] [568] [569] [570] [571] [572] [573] [574] [575] [576] [577] [578] [579] [580] [581] [582] [583] [584] [585] [586] [587] [588] [589] [590] [591] [592] [593] [594] [595] [596] [597] [598] [599] [600] [601] [602] [603] [604] [605] [606] [607] [608] [609] [610] [611] [612] [613] [614] [615] [616] [617] [618] [619] [620] [621] [622] [623] [624] [625] [626] [627] [628] [629] [630] [631] [632] [633] [634] [635] [636] [637] [638] [639] [640] [641] [642] [643] [644] [645] [646] [647] [648] [649] [650] [651] [652] [653] [654] [655] [656] [657] [658] [659] [660] [661] [662] [663] [664] [665] [666] [667] [668] [669] [670] [671] [672] [673] [674] [675] [676] [677] [678] [679] [680] [681] [682] [683] [684] [685] [686] [687] [688] [689] [690] [691] [692] [693] [694] [695] [696] [697] [698] [699] [700] [701] [702] [703] [704] [705] [706] [707] [708] [709] [710] [711] [712] [713] [714] [715] [716] [717] [718] [719] [720] [721] [722] [723] [724] [725] [726] [727] [728] [729] [730] [731] [732] [733] [734] [735] [736] [737] [738] [739] [740] [741] [742] [743] [744] [745] [746] [747] [748] [749] [750] [751] [752] [753] [754] [755] [756] [757] [758] [759] [760] [761] [762] [763] [764] [765] [766] [767] [768] [769] [770] [771] [772] [773] [774] [775] [776] [777] [778] [779] [780] [781] [782] [783] [784] [785] [786] [787] [788] [789] [790] [791] [792] [793] [794] [795] [796] [797] [798] [799] [800] [801] [802] [803] [804] [805] [806] [807] [808] [809] [810] [811] [812] [813] [814] [815] [816] [817] [818] [819] [820] [821] [822] [823] [824] [825] [826] [827] [828] [829] [830] [831] [832] [833] [834] [835] [836] [837] [838] [839] [840] [841] [842] [843] [844] [845] [846] [847] [848] [849] [850] [851] [852] [853] [854] [855] [856] [857] [858] [859] [860] [861] [862] [863] [864] [865] [866] [867] [868] [869] [870] [871] [872] [873] [874] [875] [876] [877] [878] [879] [880] [881] [882] [883] [884] [885] [886] [887] [888] [889] [890] [891] [892] [893] [894] [895] [896] [897] [898] [899] [900] [901] [902] [903] [904] [905] [906] [907] [908] [909] [910] [911] [912] [913] [914] [915] [916] [917] [918] [919] [920] [921] [922] [923] [924] [925] [926] [927] [928] [929] [930] [931] [932] [933] [934] [935] [936] [937] [938] [939] [940] [941] [942] [943] [944] [945] [946] [947] [948] [949] [950] [951] [952] [953] [954] [955] [956] [957] [958] [959] [960] [961] [962] [963] [964] [965] [966] [967] [968] [969] [970] [971] [972] [973] [974] [975] [976] [977] [978] [979] [980] [981] [982] [983] [984] [985] [986] [987] [988] [989] [990] [991] [992] [993] [994] [995] [996] [997] [998] [999] [1000] [1001] [1002] [1003] [1004] [1005] [1006] [1007] [1008] [1009] [1010] [1011] [1012] [1013] [1014] [1015] [1016] [1017] [1018] [1019] [1020] [1021] [1022] [1023] [1024] [1025] [1026] [1027] [1028] [1029] [1030] [1031] [1032] [1033] [1034] [1035] [1036] [1037] [1038] [1039] [1040] [1041] [1042] [1043] [1044] [1045] [1046] [1047] [1048] [1049] [1050] [1051] [1052] [1053] [1054] [1055] [1056] [1057] [1058] [1059] [1060] [1061] [1062] [1063] [1064] [1065] [1066] [1067] [1068] [1069] [1070] [1071] [1072] [1073] [1074] [1075] [1076] [1077] [1078] [1079] [1080] [1081] [1082] [1083] [1084] [1085] [1086] [1087] [1088] [1089] [1090] [1091] [1092] [1093] [1094] [1095] [1096] [1097] [1098] [1099] [1100] [1101] [1102] [1103] [1104] [1105] [1106] [1107] [1108] [1109] [1110] [1111] [1112] [1113] [1114] [1115] [1116] [1117] [1118] [1119] [1120] [1121] [1122] [1123] [1124] [1125] [1126] [1127] [1128] [1129] [1130] [1131] [1132] [1133] [1134] [1135] [1136] [1137] [1138] [1139] [1140] [1141] [1142] [1143] [1144] [1145] [1146] [1147] [1148] [1149] [1150] [1151] [1152] [1153] [1154] [1155] [1156] [1157] [1158] [1159] [1160] [1161] [1162] [1163] [1164] [1165] [1166] [1167] [1168] [1169] [1170] [1171] [1172] [1173] [1174] [1175] [1176] [1177] [1178] [1179] [1180] [1181] [1182] [1183] [1184] [1185] [1186] [1187] [1188] [1189] [1190] [1191] [1192] [1193] [1194] [1195] [1196] [1197] [1198] [1199] [1200] [1201] [1202] [1203] [1204] [1205] [1206] [1207] [1208] [1209] [1210] [1211] [1212] [1213] [1214] [1215] [1216] [1217] [1218] [1219] [1220] [1221] [1222] [1223] [1224] [1225] [1226] [1227] [1228] [1229] [1230] [1231] [1232] [1233] [1234] [1235] [1236] [1237] [1238] [1239] [1240] [1241] [1242] [1243] [1244] [1245] [1246] [1247] [1248] [1249] [1250] [1251] [1252] [1253] [1254] [1255] [1256] [1257] [1258] [1259] [1260] [1261] [1262] [1263] [1264] [1265] [1266] [1267] [1268] [1269] [1270] [1271] [1272] [1273] [1274] [1275] [1276] [1277] [1278] [1279] [1280] [1281] [1282] [1283] [1284] [1285] [1286] [1287] [1288] [1289] [1290] [1291] [1292] [1293] [1294] [1295] [1296] [1297] [1298] [1299] [1300] [1301] [1302] [1303] [1304] [1305] [1306] [1307] [1308] [1309] [1310] [1311] [1312] [1313] [1314] [1315] [1316] [1317] [1318] [1319] [1320] [1321] [1322] [1323] [1324] [1325] [1326] [1327] [1328] [1329] [1330] [1331] [1332] [1333] [1334] [1335] [1336] [1337] [1338] [1339] [1340] [1341] [1342] [1343] [1344] [1345] [1346] [1347] [1348] [1349] [1350] [1351] [1352] [1353] [1354] [1355] [1356] [1357] [1358] [1359] [1360] [1361] [1362] [1363] [1364] [1365] [1366] [1367] [1368] [1369] [1370] [1371] [1372] [1373] [1374] [1375] [1376] [1377] [1378] [1379] [1380] [1381] [1382] [1383] [1384] [1385] [1386] [1387] [1388] [1389] [1390] [1391] [1392] [1393] [1394] [1395] [1396] [1397] [1398] [1399] [1400] [1401] [1402] [1403] [1404] [1405] [1406] [1407] [1408] [1409] [1410] [1411] [1412] [1413] [1414] [1415] [1416] [1417] [1418] [1419] [1420] [1421] [1422] [1423] [1424] [1425] [1426] [1427] [1428] [1429] [1430] [1431] [1432] [1433] [1434] [1435] [1436] [1437] [1438] [1439] [1440] [1441] [1442] [1443] [1444] [1445] [1446] [1447] [1448] [1449] [1450] [1451] [1452] [1453] [1454] [1455] [1456] [1457] [1458] [1459] [1460] [1461] [1462] [1463] [1464] [1465] [1466] [1467] [1468] [1469] [1470] [1471] [1472] [1473] [1474] [1475] [1476] [1477] [1478] [1479] [1480] [1481] [1482] [1483] [1484] [1485] [1486] [1487] [1488] [1489] [1490] [1491] [1492] [1493] [1494] [1495] [1496] [1497] [1498] [1499] [1500] [1501] [1502] [1503] [1504] [1505] [1506] [1507] [1508] [1509] [1510] [1511] [1512] [1513] [1514] [1515] [1516] [1517] [1518] [1519] [1520] [1521] [1522] [1523] [1524] [1525] [1526] [1527] [1528] [1529] [1530] [1531] [1532] [1533] [1534] [1535] [1536] [1537] [1538] [1539] [1540] [1541] [1542] [1543] [1544] [1545] [1546] [1547] [1548] [1549] [1550] [1551] [1552] [1553] [1554] [1555] [1556] [1557] [1558] [1559] [1560] [1561] [1562] [1563] [1564] [1565] [1566] [1567] [1568] [1569] [1570] [1571] [1572] [1573] [1574] [1575] [1576] [1577] [1578] [1579] [1580] [1581] [1582] [1583] [1584] [1585] [1586] [1587] [1588] [1589] [1590] [1591] [1592] [1593] [1594] [1595] [1596] [1597] [1598] [1599] [1600] [1601] [1602] [1603] [1604] [1605] [1606] [1607] [1608] [1609] [1610] [1611] [1612] [1613] [1614] [1615] [1616] [1617] [1618] [1619] [1620] [1621] [1622] [1623] [1624] [1625] [1626] [1627] [1628] [1629] [1630] [1631] [1632] [1633] [1634] [1635] [1636] [1637] [1638] [1639] [1640] [1641] [1642] [1643] [1644] [1645] [1646] [1647] [1648] [1649] [1650] [1651] [1652] [1653] [1654] [1655] [1656] [1657] [1658] [1659] [1660] [1661] [1662] [1663] [1664] [1665] [1666] [1667] [1668] [1669] [1670] [1671] [1672] [1673] [1674] [1675] [1676] [1677] [1678] [1679] [1680] [1681] [1682] [1683] [1684] [1685] [1686] [1687] [1688] [1689] [1690] [1691] [1692] [1693] [1694] [1695] [1696] [1697] [1698] [1699] [1700] [1701] [1702] [1703] [1704] [1705] [1706] [1707] [1708] [1709] [1710] [1711] [1712] [1713] [1714] [1715] [1716] [1717] [1718] [1719] [1720] [1721] [1722] [1723] [1724] [1725] [1726] [1727] [1728] [1729] [1730] [1731] [1732] [1733] [1734] [1735] [1736] [1737] [1738] [1739] [1740] [1741] [1742] [1743] [1744] [1745] [1746] [1747] [1748] [1749] [1750] [1751] [1752] [1753] [1754] [1755] [1756] [1757] [1758] [1759] [1760] [1761] [1762] [1763] [1764] [1765] [1766] [1767] [1768] [1769] [1770] [1771] [1772] [1773] [1774] [1775] [1776] [1777] [1778] [1779] [1780] [1781] [1782] [1783] [1784] [1785] [1786] [1787] [1788] [1789] [1790] [1791] [1792] [1793] [1794] [1795] [1796] [1797] [1798] [1799] [1800] [1801] [1802] [1803] [1804] [1805] [1806] [1807] [1808] [1809] [1810] [1811] [1812] [1813] [1814] [1815] [1816] [1817] [1818] [1819] [1820] [1821] [1822] [1823] [1824] [1825] [1826] [1827] [1828] [1829] [1830] [1831] [1832] [1833] [1834] [1835] [1836] [1837] [1838] [1839] [1840] [1841] [1842] [1843] [1844] [1845] [1846] [1847] [1848] [1849] [1850] [1851] [1852] [1853] [1854] [1855] [1856] [1857] [1858] [1859] [1860] [1861] [1862] [1863] [1864] [1865] [1866] [1867] [1868] [1869] [1870] [1871] [1872] [1873] [1874] [1875] [1876] [1877] [1878] [1879] [1880] [1881] [1882] [1883] [1884] [1885] [1886] [1887] [1888] [1889] [1890] [1891] [1892] [1893] [1894] [1895] [1896] [1897] [1898] [1899] [1900] [1901] [1902] [1903] [1904] [1905] [1906] [1907] [1908] [1909] [1910] [1911] [1912] [1913] [1914] [1915] [1916] [1917] [1918] [1919] [1920] [1921] [1922] [1923] [1924] [1925] [1926] [1927] [1928] [1929] [1930] [1931] [1932] [1933] [1934] [1935] [1936] [1937] [1938] [1939] [1940] [1941] [1942] [1943] [1944] [1945] [1946] [1947] [1948] [1949] [1950] [1951] [1952] [1953] [1954] [1955] [1956] [1957] [1958] [1959] [1960] [1961] [1962] [1963] [1964] [1965] [1966] [1967] [1968] [1969] [1970] [1971] [1972] [1973] [1974] [1975] [1976] [1977] [1978] [1979] [1980] [1981] [1982] [1983] [1984] [1985] [1986] [1987] [1988] [1989] [1990] [1991] [1992] [1993] [1994] [1995] [1996] [1997] [1998] [1999] [2000] [2001] [2002] [2003] [2004] [2005] [2006] [2007] [2008] [2009] [2010] [2011] [2012] [2013] [2014] [2015] [2016] [2017] [2018] [2019] [2020] [2021] [2022] [2023] [2024] [2025] [2026] [2027] [2028] [2029] [2030] [2031] [2032] [2033] [2034] [2035] [2036] [2037] [2038] [2039] [2040] [2041] [2042] [2043] [2044] [2045] [2046] [2047] [2048] [2049] [2050] [2051] [2052] [2053] [2054] [2055] [2056] [2057] [2058] [2059] [2060] [2061] [2062] [2063] [2064] [2065] [2066] [2067] [2068] [2069] [2070] [2071] [2072] [2073] [2074] [2075] [2076] [2077] [2078] [2079] [2080] [2081] [2082] [2083] [2084] [2085] [2086] [2087] [2088] [2089] [2090] [2091] [2092] [2093] [2094] [2095] [2096] [2097] [2098] [2099] [2100] [2101] [2102] [2103] [2104] [2105] [2106] [2107] [2108] [2109] [2110] [2111] [2112] [2113] [2114] [2115] [2116] [2117] [2118] [2119] [2120] [2121] [2122] [2123] [2124] [2125] [2126] [2127] [2128] [2129] [2130] [2131] [2132] [2133] [2134] [2135] [2136] [2137] [2138] [2139] [2140] [2141] [2142] [2143] [2144] [2145] [2146] [2147] [2148] [2149] [2150] [2151] [2152] [2153] [2154] [2155] [2156] [2157] [2158] [2159] [2160] [2161] [2162] [2163] [2164]

21. Jernigan RL & Sauer SA (1989) Structure of T4 DNA polymerase complexed to the transcription factor T7 polymerase. *EMBO J* 11: 3921-3923.
22. Zouze R, Sauer SA & Wang H (1999) The polymerase complexed to the transcription factor T7 polymerase is present in the structure of a crystal. *J Biol Chem* 274: 8112-8116.
23. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
24. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
25. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
26. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
27. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
28. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
29. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
30. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
31. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
32. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
33. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
34. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
35. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
36. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
37. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
38. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
39. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
40. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
41. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
42. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
43. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
44. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
45. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
46. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
47. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
48. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
49. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.
50. Jernigan RL, Sauer SA, Wang H, Sauer SA & Sauer SA (1999) Crystal structure of the transcription factor T7 polymerase. *Science* 285: 812-816.

Because *Staphylococcus aureus* is a pathogenic bacterium, this paper has been published in the *Journal of Bacteriology* (published from <http://journals.asm.org/>). For further information, see the explanation on the customer page.

# Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*

(X-ray structure/double bond/crystallization/sequence/thermostable)

KARL-PETER HOPFNER\*<sup>†</sup>, ANDREAS EICHINGER\*, RICHARD A. EGIN\*<sup>‡</sup>, FRANK LAINE<sup>§</sup>, WALTER ANGENBACHER<sup>§</sup>, ROBERT FRIDER\*, AND BERNHARD AMERER<sup>§</sup>

<sup>\*</sup>Abteilung Biomolekularbiologie, Max-Planck-Gesellschaft, D-80335 München, Germany, and <sup>§</sup>Strukturbiologisches Institut, Universität Hamburg, Germany

Contributed by Robert Frider, January 22, 1999

**ABSTRACT** Most known archaeal DNA polymerases belong to the type B family, which also includes the DNA replication polymerases of eukaryotes, but maintain high fidelity at extreme conditions. We describe here the 2.5 Å resolution crystal structure of a DNA polymerase from the archaeon *Thermococcus gorgonarius* and identify structural features of the fold and the active site that are likely responsible for its thermostable function. Comparison with the mesophilic B type DNA polymerase gp43 of the bacteriophage T4 highlights thermophilic adaptations, which include the presence of two disulfide bonds and an enhanced electrostatic complementarity in the DNA-protein interface. In contrast to gp43, several loops in the exonuclease and thumb domains are more closely packed; this apparently blocks primer binding to the exonuclease active site. A physiological role of this "closed" conformation is unknown but may represent a polymerase switch, in contrast to an editing mode with an open exonuclease site. This archaeal B DNA polymerase structure provides a starting point for structure-based design of polymerases or ligands with applications in biotechnology and the development of anticancer or anticancer agents.

Propagation of DNA requires faithful DNA replication. This is performed in vivo by DNA polymerase (pol), which attaches the appropriate dNTP to the nascent DNA primer strand in match its paired template. Different families of pols are involved in different DNA polymerization processes including not only DNA replication (1, 2) but also repair and recombination (3, 4), a heterogeneity also reflected by varying polypeptide structures and/or catalytic compositions (5, 6). Some pols complement polymerase activity with 3' → 5' exonuclease activity (editing activity) and/or 5' → 3' terminal transferase (nickase) activity, often located in separate structural domains on the same polypeptide chain (4–8).

Crystal structures are available for most known polymerase families, including the A family DNA polymerases (9–14), pol  $\beta$  (15–17), HIV reverse transcriptase (18–20), and, recently, the B family pol gp43 from bacteriophage T4 (21). All share a functional polymerase structure, which resembles a right hand built by the palm, finger and thumb domains (see ref. 7 for review). Although the fingers and thumb domains are highly diverse among the different families, the palm domains, which contain the conserved catalytic aspartate residues, show a similar topology among all families except pol  $\beta$ . The polymerase nucleotide transfer was studied in detail for the A family polymerases, HIV reverse transcriptase, and pol  $\beta$ , and was shown to involve two metal ions (summarized in ref. 7).

Considerably less is known for the family of type B pols, which are replicative enzymes in eukaryotes and most likely

also Archaea (22, 23). The structure of gp43 from bacteriophage T4 (21) provided an excellent first insight into this family. In addition to the three polymerase domains, gp43 contains an 3' → 5' exonuclease domain and an 5' terminal domain. The exonuclease and palm domains share the topology and active site of A family enzymes, implying similar mechanistic mechanisms for polymerase and exonuclease activities (24). The thumb and finger domains are apparently unrelated to the other polymerase families. The function of the 5' terminal domain remains unknown, but may help assemble the multicomponent replication apparatus (24).

Much is known about the replication of plasmids (24–26), viruses (1, 27), Penicillium (28) and Eukaryotes (1, 2, 29, 30), which in general involves pols that also possess nickase, helicase, RNaseH, binding cleavage, and other factors (31). Considerably less is known for archaeal replication, which mostly B type polymerases, similar to eukaryotic replication enzymes pol  $\alpha$  and  $\delta$ , have been identified (6, 22, 23, 32–34). This relative ignorance is surprising, because such crucial biotechnological applications as cloning and PCR require the thermostability and fidelity typical of archaeal polymerases (4). Thus, in addition to satisfying basic research interests, structural information could assist, for example, the engineering of variant enzymes with tailored nucleotide incorporation rates or the design of antisense and anticancer polymerase inhibitors. For these reasons, we have determined the structure of a DNA polymerase from *Thermococcus gorgonarius* (Tgo), an extremely thermophilic sulfur metabolizing archaeon isolated from a geothermal vent in New Zealand (35). This enzyme possesses pol and a 3' → 5' exonuclease activity, which together ensure thermostable replication with high fidelity (turnover: 3.2–5.2 × 10<sup>4</sup> sec<sup>-1</sup>; see ref. 36). The 2.5 Å structure shows a topological similarity to gp43 and gives insight into the structural biology of archaeal DNA polymerases, including the identification of several mechanisms for thermostable adaptation.

## MATERIALS AND METHODS

**Materials.** All materials were of the highest grade commercially available.

**Bacterial Strains.** *Escherichia coli* L33392 containing pT455.26 was used as described (36). *E. coli* HB34 (DE3) (this study) was a generous gift of Matthias Bode (Max-Planck-Institut).

**Expression Vectors.** pBT222 was obtained from Roche Molecular Biochemicals.

**Abbreviations:** Tgo, *Thermococcus gorgonarius* pol; DNA polymerase; DNA dependent. The amino acid sequence has been deposited in the Protein Data Bank, Biology Department, Brookhaven National Laboratory, Upton, NY, 11973 (PDB ID code 1TGO). Present address: The Scripps Research Institute, La Jolla, CA 92037. To whom reprint requests should be addressed; e-mail: hopfner@scripps.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 16 U.S.C. §1754 only to indicate this fact.

© 1999 by Academic Press. 0896-6460/99/035682-05\$05.00/0

5682

Table 1. Data collection and isomorphous replacement

Derivative	Resolution limit, Å	Total observations	Unique observations	Completeness, %	R <sub>int</sub> , %	R <sub>meas</sub> , %	Phasing power
Native	3.0	136,973	21,479	92.9	7.9		
U	3.2	25,994	16,158	66.5	11.9	18.5	0.71
PT1	3.7	65,484	13,893	56.2	12.6	22.6	1.54
PT2	4.0	53,670	6,291	66.9	13.5	13.8	1.77
PT3	3.4	85,726	14,153	92.8	16.5	17.1	1.16
PT4	3.6	387,922	27,487	90.6	3.7	15.3	0.67
PT5	2.7	320,695	24,543	81.7	6.1	16.2	0.51
PT12	5.5	59,629	12,477	82.4	9.8	16.7	1.54
PD	5.5	78,136	11,336	92.9	11.5	12.8	0.42
PBPT	5.4	64,557	14,129	92.6	8.9	18.8	1.04
OS	3.8	698,168	24,926	91.2	5.2	18.4	1.82

Overall figure of merit (15.0–17.5 Å): 0.73

Heavy-atom derivatives were prepared by soaking the crystals in low salt buffer containing the heavy atom solution: U, 0.5 mM uranyl acetate for 2 h; PT1, 5 mM K<sub>2</sub>PtCl<sub>6</sub> for 1 d; PT2, 5 mM K<sub>2</sub>PtCl<sub>6</sub> for 2 d; PT3, saturated cadmium chloride (CdCl<sub>2</sub>) for 3 d; PT4, 5 mM K<sub>2</sub>PtCl<sub>6</sub> for 7 d; PT5, saturated AgNO<sub>3</sub> for 2 d; PT12, PT3 + U; PD, saturated diiodomaleonitrile (DIO) for 3 d; PBPT, PD + PT5; OS, saturated K<sub>2</sub>SO<sub>4</sub> for 7 d. NAT1, PT1, PT2, PT3, PT4, PD and PBPT were collected with a Mar imaging plate and U was collected with a Bruker AXS area detector on a Rigaku rotating anode beam. All other data sets were collected with a Mar image-coupled device at beamline BW6 at DESY, Hamburg.

**Cloning, Expression, and Protein Purification.** The gene for the 89.8 kDa DNA-dependent pol (Tpo pol) was cloned from Tpo (Clostridia) genomic DNA using primers 5'-GGTGG and 3'-GGTGG and expressed in *E. coli* LE392/pL108320 plasmid (Dietrich) using the M13mp18 expression vector (13,32) as described (36). Tpo pol was purified essentially as described (36) with the addition of the TSK, butyl Toyopearl column by Blue-Trans Sepharose (Bio-rad) and with an additional desalting step on Pharm SE HPLC using exchange medium (Boehringer Mannheim Biochemicals). Active fractions were combined, concentrated to 30 mg/ml, and transferred to 20 mM sodium phosphate, pH 8.2/10 mM 2-mercaptoethanol/500 mM NaCl.

The gene for a salmonella-like containing variant of Tpo pol (Se-Tpo pol) was expressed in *E. coli* B942 (DE3) (Pharmacia) using a published protocol (37). Se-Tpo pol was purified by using the wild-type protocol.

**Crystallization.** Crystals of purified Tpo pol (for Se-Tpo pol) were grown by using sitting drop vapor diffusion technique at 25°C with high salt conditions (5.2 M potassium acetate/100 mM Tris, pH 8.0/0.8 M ammonium sulfate) and subjected to 3.0 Å (high salt) and to 2.5 Å (low salt) beamline BW6 at Deutsches Elektronen-Synchrotron (DESY, Hamburg). Low salt conditions (100 mM Tris, pH 7.0/2.0 M ammonium sulfate/10% PEG 400) yielded only poorly diffracting crystals but allowed soaking (including heavy atoms) with some cell constant modifications (*a*, *b*, *c* = 63.6, 105.0, 190.5 Å) but minimal loss of resolution.

**Data Collection and Processing.** Data were collected with a MAR imaging plate or a Bruker AXS KODAK mounted on a Rigaku rotating anode source, or with a MAR imaging plate or a MAR CCD (charge-coupled device) at beamline BW6 at DESY, Hamburg. The data were processed with Bruker AXS, version 4 (Mar CCD) or 38.0 or 10800 (MAR imaging plate) or 36.0, scaled with SCALE (40) or SCALEPACK (39), and reduced with TRUNCATE (40).

**Structure Determination.** The structure was solved by multiple isomorphous replacement and anomalous scattering (MIRAS) by using data from crystals transferred to low salt conditions (Table 1). Crystallographic calculations were done with programs from the CCP4 suite (40). Heavy atom positions of major sites were located in difference Patterson maps and were refined with OUTLINE (40) in calculated profile phase angles to 3.5 Å resolution. A partial polyatomic model was built into interpretable positions of secondary structural elements of the initial map by using MAIN (41). The quality of the electron density was improved by phase combination of the partial model with the experimental phases by using ROTATE

(40), and several cycles of solvent flattening to 3.0 Å by using SOLVENT (40). At this stage, no interpretable density was found for a significant portion of the molecule, comprising residues 147–154, 263–266, 653–728, and 752–773.

**Model Building and Refinement.** The partial model (8 factor 35%) was used to phase the 2.5 Å resolution data of the Se-Tpo pol (high salt conditions). The model was oriented with ASCOR (40). The correlation coefficient of 22.0% and the R factor of 50.3% showed divergence of the high- and low-salt structures. After bulk solvent correction, anisotropic B factor correction and rigid-body minimization (treating five domains independently), the partial model was iteratively refined and extended with simulated annealing, Powell minimization, restrained individual B factor refinement with CNS (42), and manual model building with MAIN (41) by using data from 2.5–2.5 Å resolution (Table 2, Fig. 1).

## RESULTS AND DISCUSSION

**Structure of Tpo pol.** Tpo pol is a long, compact molecule with dimensions 50 Å × 80 Å × 150 Å. The single polypeptide chain of 773 aa is folded into five distinct structural domains (Fig. 2): the N-terminal domain (residues 1–130), the 3' → 5' exonuclease domain (131–520), the palm (369–449) and 3'-5' endo- (450–489), and thumb (530–773) domains of the polymerase unit, and a helical interdomain-interaction (527–580).

Table 2. Crystallographic refinement, Rigaku data

Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions, Å	<i>a</i> = 58.1, <i>b</i> = 105.2, <i>c</i> = 154.2
Observations, 2θ–2.5 Å	
Total	421,449
Unique	30,431
Completeness, %	
Total	92.1
Last shell	66.6
R <sub>int</sub> , %	
Total	7.2
Last shell	20.2
R factor (R <sub>int</sub> ), %	20.9 (22.5)
excl deviation to bond lengths, Å	0.006
excl deviation to bond angles, °	1.5
No. of nonhydrogen atoms	
Protein	6,778
Water	330



Fig. 1. Stereorepresentation of the electron-density map. The  $2(F_o - F_c)$  electron density (contoured at 1.5  $\sigma$ ) at the polymerase active site is well defined for the ribbon model (stick representation).

between the exonuclease and palm domains. The polymerase unit forms the DNA-binding clefts, reminiscent of a right hand, which is the identifying characteristic of pols. Gp43 from bacteriophage RB69 also shows this overall domain topology (21).

Three clefts extend radially from the polymerase active site at the center of the ring; two of them in opposite directions, forming a large cleft across the molecule, and one perpendicular to these. Based on active-site homology to pol A family enzymes, the two opposite clefts probably bind duplex DNA (cleft D, according to ref. 21) and single-strand template DNA (cleft T), respectively. The perpendicular (editing) cleft links

the polymerase active site and the exonuclease active site and binds the primer strand in editing mode (21).

The exonuclease domain is structurally equivalent to the 3'  $\rightarrow$  5' exonuclease domain of pol A family (49). Like gp43, however, it is bound at the opposite side of the polymerase unit by noncovalent contacts to the thumb domain at the editing cleft, on one side, and by covalent and noncovalent contacts to the 3'-terminal and palm domains and the  $\alpha 2$  residues inter-domain helix, on the other side. This latter segment is located at the base of cleft T, which is additionally bounded by the exonuclease, 3'-terminal, and palm domains.

The topology of the palm domain is conserved among polymerase families (5), with two long helices (G and R)

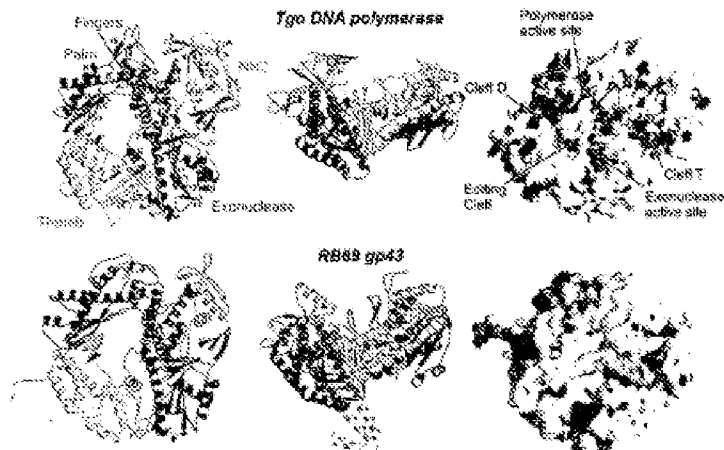


Fig. 2. Structure of Tgo pol and comparison with gp43 from bacteriophage RB69. (Left and middle) Ribbon representation of Tgo DNA polymerase (Upper) in two orthogonal views with selected secondary structure elements. The molecule is composed of five domains. N-terminal domain (yellow), 3'  $\rightarrow$  5' exonuclease (red), palm (light and dark orange) represent the A domain and C-terminus segment, respectively (see text). Finger (blue), and thumb (green), which are arranged to form a ring. An interdomain helical segment between the exonuclease domain and the palm is orange. The conserved substructure in the active site and the two double-strand binding sites are shown as orange and yellow rods and sticks, respectively. Tgo pol has the same overall architecture and domain topology than gp43 of RB69 (Lower). The N-terminal domain in the finger of gp43 is gray. (Right) Comparison of molecular surfaces of Tgo pol (Upper) and RB69 gp43 (Lower). Red and blue denote negative and positive electrostatic surface potentials, respectively. In contrast to gp43, Tgo pol has a strongly extended positive potential at the positive DNA-binding cleft.

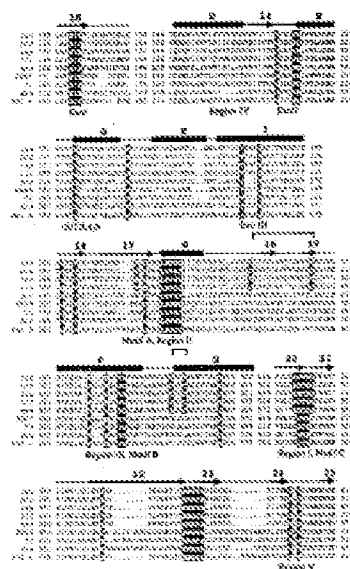


Fig. 3. Sequence alignment of B family DNA polymerases. The alignment has been adapted from ref. 21 to highlight specific residues from the class of archaeal pols. The secondary structure of T4 pol is indicated at top of the alignment with helices (black), loops (gray), and single turns (white) colored according to domains with the same color code as Fig. 2B. Strikingly conserved residues of type B polymerases are red, and additional conserved residues are yellow. Uniquely conserved residues of archaeal type B enzymes are indicated in the less-conserved green. Disulfide bonds are shown by a bar on top of the alignment. Abbreviations: *geo*, *Thermococcus geophilus* pol; *gla*, *Pyrococcus furiosus* pol; *tp*, *Thermoplasma* sp. pol; *tr*, *Thermococcus thermophilus* pol; *ova*, *Methanococcus voltae* pol; *R689*, bacteriophage R689 pol; *T4*, bacteriophage T4 pol; *E. coli*, *E. coli* pol; *h*, *h* pol; *h*, *h* pol; *h*, *h* pol.

packed against the five-stranded antiparallel  $\beta$ -sheet that underlies the three conserved aspartate residues involved in nucleoside transfer. The finger emerges from the nucleic acid channel as a helix with insertion. Its 20 residues are folded into two independent  $\alpha$ -helices of approximately equal size; helix P contains the conserved K484VY485 motif of B type polymerases and is related to the C helix of A type enzymes (see below). The ~60 residue insertion between helix P and P in R689 and T4 gp43 is missing in T4 pol, where both helices and a residue linker are much shorter than their equivalents in gp43. The shorter fingers of T4 pol presumably reflect the typical structure of the nonhomologous B type fingers (see, e.g., Fig. 4 and 5, and 6, and 7). The thumb domain topology, identical to that of gp43, is unrelated to other polymerase types. However, like the thumb of A type enzymes, a bundle of  $\alpha$ -helices at its base protrudes from the active site  $\beta$ -sheet flaps at the active site, the thumb contains a 78-residue subdomain (665–729), which faces the exonuclease domain and

contributes to the sliding channel, explaining why mutations in the exonuclease domain of B-type polymerases affect the polymerase activity and vice versa (44, 45).

Weakly defined density increases the base of the thumb domain was indicated as the C-terminal  $\beta$ -residue with a polyanionic chain. The C-terminal  $\beta$ -residue does not protrude from the active site as in the R689 polymerase (21). Because the C terminus of the T4 pol is involved in sliding channel binding (46), it is likely, however, that these residues become ordered to any similar holonucleic formation.

**Sequence Alignment of Archaeal DNA Polymerases.** The structure of T4 pol allows the generation of a structure-based sequence alignment of the archaeal subfamily of type B DNA polymerases, the location of conserved and unique residues, and the comparison with other type B DNA polymerases (Fig. 3).

**Polymerase Active Site.** The polymerase active site is formed by the central  $\beta$ -sheet strands 16, 17, 20, 21, and 22 and helix N4 of the palm domain and helix P located in the finger and is highly conserved among B family polymerases (Fig. 4). Three carboxylates required for nucleoside transfer in B family polymerases, two of which coordinate two metal ions (34) are superimposedly conserved among A family enzymes, B family enzymes, and reverse transcriptase (23). Superposition of T4 pol and T7 replicase complex (14) places the dNTP near the proposed nucleoside-binding site in helix P, the K484VY485 motif (Fig. 3) and suggests interactions of the carboxylate with metal and the phosphate end of the bound dNTP (Fig. 4). Superposition of the strictly conserved Tyr-487 allows it to mimic the Tyr-522-phosphate salt interaction in T7, Tyr-484 (K484VY485 motif) and Tyr-489 (SLN-489PSE) from the bottom of the nucleoside binding site.

The active site of B family polymerases contains a DTDS motif, which, however, is DTDC in the archaeal subfamily. In T4 pol, the relatively conserved Tyr-482 from the adjacent strand provides an electron-shielded group, it is positioned appropriate for metal coordination or binding of the 3' end of the primer. The orientation of Tyr-482 is stabilized by an aromatic cluster that also includes Phe-543 and Tyr-538.

Archaeal *Methanococcus voltae* and *Thermococcus* sp. gp43 (see Fig. 3) have Tyr at position 543—Phe in T4 pol—rather than at 482, but might also carry an electron group. The displacement of a functional cluster from active site in DTDS to Tyr-482 or Tyr-543 might stabilize its orientation as an adaptation for thermotolerance.

The conserved cluster of acidic amino acids (557S, E560) forms an unexpected metal-binding site for  $Mg^{2+}$  and  $Zn^{2+}$  (Fig. 4). Its proximity to Asp-494 and to the expected location of the dNTP  $\gamma$ -phosphate suggests a supporting role in nucleoside binding and/or catalysis.

**3'  $\rightarrow$  5' Exonuclease Active Site.** T4 pol is characterized by a strong 3'  $\rightarrow$  5' exonuclease activity, unlike eukaryotic B type polymerases (unpublished results). The exonuclease active site is formed at the interface between the exonuclease domain and the  $\beta$  of the thumb (Fig. 5). All residues required for nucleoside transfer are located in the exonuclease domain, which, at least for T4 gp43, retains activity when dissociated from the polymerase (47). However, the thumb domain, with, for example, R689 gp43's Phe-123-base intercalation, partially controls the binding geometry of single strand DNA (21, 45).

The exonuclease structures of T4 and gp43 DNA polymerases are similar in the sliding site but differ considerably at the exonuclease domain interface. Struts 19 contains the metal-binding D181E motif and readily superimposes with the equivalent strand from gp43, allowing modeling of a single-strand DNA segment into the exonuclease site (see Fig. 5 and 6). The conserved residues Asp-181 and Glu-143 in the Eco I motif, Tyr-389, Asp-410, Phe-254, and Asp-215 in Eco II, and Tyr-311 and Asp-315 in Eco III are in approximate DNA-binding nonconsensus (Fig. 5).

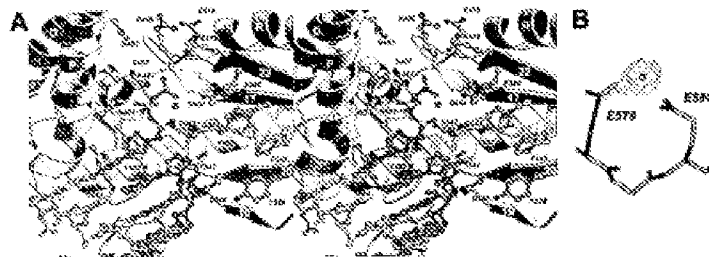


Fig. 4. Polymerase active site. (A) Stereorepresentation (using color code as in Fig. 2) with residues 186A. Across the residues are shown as ball-and-stick representations with carbon (green), nitrogen (blue), and oxygen (red) atoms. The 186A, 186B, 186C, 186D, 186E, 186F, 186G, 186H, 186I, 186J, 186K, 186L, 186M, 186N, 186O, 186P, 186Q, 186R, 186S, 186T, 186U, 186V, 186W, 186X, 186Y, 186Z, and adjacent residues with corresponding residues in T7 pol (D412 and D444). Phosphorus atoms are yellow. The conformation of the T7 replication complex is shown as magenta spheres. (B) Experimentally observed metal-binding site for  $Mg^{2+}$ ,  $Mn^{2+}$ ,  $Co^{2+}$ ,  $Ni^{2+}$ ,  $Zn^{2+}$ , and  $Cd^{2+}$  in the "low salt" crystal form. The residues E376 and E380 are conserved in type B polymerases (Fig. 2).

5A). However, the sliding shift is constrained by a displacement of the tip of the thumb toward the exonuclease domain to prohibit single-strand binding (Fig. 5B). This shift is correlated with a large change in the loop between strands 18 and 19, to RDE8 gp43 (and likewise in T4 gp43), that loop forms a bulge near the 3' base and contains Phe-173, which intercalates between the first two bases. In T7 pol, the loop is oriented outward, away from the thumb, and Phe-152 (the equivalent of gp43's Phe-173) straddles the Phe-214 10 Å away from intercalation site. This shift allows the tip of the thumb to move into the editing channel and to block the exonuclease site.

**Are There Different Conformations in Polymerase and Editing Mode?** If a closed conformation of the exonuclease domain prohibits single-strand binding, an open conformation is required for editing. The observed closed conformation may represent the enzyme in "polymerase" mode. Preliminary analysis of the crystal structure of T7 pol in the low-salt condition indicates a structural change at the interface of exonuclease and thumb, possibly reflecting a transition be-

tween open and closed states. The closed conformation observed here may, however, be a nonphysiological artifact of the high ionic strength used for crystallization. Crystal structures of the enzyme in both polymerase and editing modes are required.

**Adaptation to High Temperatures.** T7 pol is a sulfur-metabolizing, extremely thermophilic archaeon, with a growth range between 55°C and 80°C. For accurate replication at this temperature range, the polymerase must not only be stable, but must also adequately bind substrate DNA. A comparison with gp43 from the mesophilic bacteriophage  $\phi$ 29 indicates several such adaptations to high temperatures. Several loops are shorter in T7 pol than in gp43 (Fig. 2), and there is an increase in hydrogen-bonded  $\beta$ -strand content: T7 pol secondary structure includes 43%  $\alpha$ -helix, 22%  $\beta$ -strands, and 19% turns (calculated according to ref. 30), whereas gp43 has 42%  $\alpha$ -helix, 17%  $\beta$ -strands, and 19% turns.

Although rare among cytoplasmic or nuclear proteins, two double-strand breaks occur in T7 pol: cysteine pairs 428–442 and



Fig. 5. 5' to 3' exonuclease active site. (A) Stereorepresentation with residues 186A, 186B, 186C, 186D, 186E, 186F, 186G, 186H, 186I, 186J, 186K, 186L, 186M, 186N, 186O, 186P, 186Q, 186R, 186S, 186T, 186U, 186V, 186W, 186X, 186Y, 186Z, and Fig. 2 (ribbons). Across the residues are shown as ball-and-stick representation. The single-stranded DNA (ssDNA) is shown from the conformation of RDE8-gp43 complex (21). The conformation of the DNA has been obtained by superimposing the EDE343 motif in T7 pol with corresponding motif of RDE8 gp43 (DNE116). Strand 27 and its preceding loop from the thumb (green) is superimposed in collision with the modified DNA. (B) Conformation of the exonuclease-thumb interface between T7 pol (color code of Fig. 2B) and RDE8 gp43 (gray). In T7 pol, the tip of the editing site (red) is here oriented toward the equivalent loop of gp43 (yellow), allowing the tip of the thumb to move toward A (4) close to the exonuclease. This conformation is incompatible with formation of an editing complex (the point) of gp43 shown as brown open-shell model.

We thank Dr. Tadeo Jacci and Néstor Agustín for stimulating discussions, Dr. Mediljo Budija for help with expression of solenome spinning variants, and Dr. Hans Reineck and Dr. Gert Gonsky of DDTV, Harbours for help with data collection.

1. Waga, S. & Saitoh, R. (1989) *Acta Soc. Sci. Ser. Biomed.* **47**, 721-729.
2. Waga, S. & Saitoh, R. (1990) *Bio. Sci.* **55**, 3354-3377.
3. Wang, T. S. F. (1991) *Ann. Rev. Biochem.* **60**, 513-552.
4. Kneebone, A. & Baker, T. A. (1992) *Local Anesthetics* (Oxford Science Publications).
5. Jones, C. & Saito, T. A. (1994) *Ann. Rev. Biochem.* **63**, 773-822.
6. Verker, F. S., Knaus, S. & Kung, H. (1986) *Adv. Protein Chem.* **41**, 1-50.
7. Steinberg, C. A. & Smith, T. A. (1988) *Chem. Opin. Struct. Biol.* **3**, 32-33.
8. Arnold, E., Fong, Z., Heggie, S. & deWitt, R. (1995) *Chem. Opin. Struct. Biol.* **5**, 27-35.
9. Li, D. L., Breck, N., Harlow, R., Xiang, N. Q. & Tsao, T. A. (1993) *Nature* **362**, 351-353.
10. Freeman, P. S., Freeman, J., Deane, S. S., Sorenson, M. S. & Tsao, T. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 1234-1238.
11. Kierulff, S. M., Moss, C., Horrocks, C. T., Svedas, S. L., Bergsht, H. H., Shannin, R. C. & Brown, T. S. (1997) *Nature* **387**, 65-68.
12. Kierulff, S. M., Moss, C., Brown, T. C. & Brown, T. S. (1998) *Nature* **394**, 384-386.
13. Kuo, Y., Rini, S. H., Wang, J., Lee, D. S., Rob, S. W. & Steitz, T. A. (1995) *Nature* **376**, 612-618.
14. Doolittle, S., Tobin, S., Wang, J., Kierulff, S. M., Steinberg, C. A. & Tsao, T. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 221-223.

- [illegible]



#### Appendix I

We have purified and characterized the Family B DNA polymerase from the archaeon *Methanococcus maripaludis*, cloned from ATCC 43090. This polymerase has a 41% sequence identity and 63% sequence similarity with Vent DNA Polymerase when analyzed using NCBI Blast 2 and the default parameters.

We performed the titration assay described in Example 1 of the patent application, using the Mma, Vent (exo-), and 9°N (exo+) DNA Polymerases. Experimental details and data are given in the attached figure.

For each of the three polymerases, a comparison of lanes using dideoxycyCTP (ddCTP) with those using equivalent concentrations of acycloCTP (acyCTP) reveals shorter products in lanes utilizing acyCTP. These shorter products result from more efficient insertion of the acyCTP terminator compared to incorporation of the ddCTP terminator. Thus, all three polymerases incorporated acyCTP more efficiently than ddCTP.

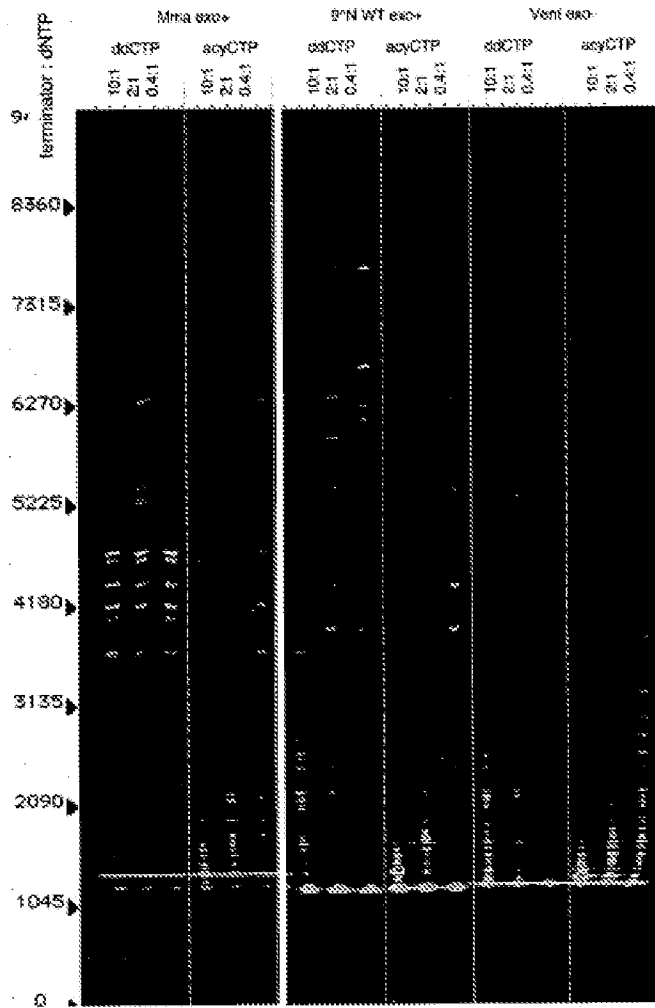
#### Figure Legend

The ability of acyNTPs and ddNTPs to act as chain terminators was tested using a titration assay of the type described in Example 1. Incorporation of ddCTP was compared to that of acyCTP, respectively, using *Methanococcus maripaludis* DNA polymerase, 9°N (exo+) DNA polymerase and Vent® (exo-) DNA polymerases.

Incorporation of ddCTP and acyCTP was assayed by mixing 8 µl of reaction cocktail (0.025 µM 5' [FAM] end-labeled #1224-primed M13mp18, 62.5 mM NaCl, 12.5 mM Tris-HCl (pH 7.9 at 25°C), 12.5 mM MgCl<sub>2</sub>, 1.25 mM

dithiothreitol, *Methanococcus maripaludis* DNA polymerase or 0.125 U/ $\mu$ l 9°N (exo+) DNA polymerase or 0.125 U/ $\mu$ l Vent® (exo-) DNA polymerase) with 2  $\mu$ l of 5X nucleotide analog/nucleotide solution to yield the final ratios of analog:dNTP indicated in the figures. After incubating at 72°C for 20 minutes, the reactions were halted by the addition of 10  $\mu$ l formamide. Samples were then heated at 72°C for 3 minutes and a 1  $\mu$ l aliquot was loaded on a 4% polyacrylamide urea gel and detected by an ABI377 automated DNA sequencer.

ddCTP v. acyCTP incorporation by archaeal DNAPs



BEST AVAILABLE COPY